



# BioExtract Server

## User Manual

**University of South Dakota**

**About Us** The BioExtract Server harnesses the power of online informatics tools for creating and customizing workflows. Users can query online sequence data, analyze it using an array of informatics tools (web service and desktop), create and share custom workflows for repeated analysis, and save the resulting data and workflows in standardized reports. This work was initially supported by NSF [grant 0090732](#). Current work is being supported by NSF [DBI-0606909](#).

Copyright © 2008 [Brendel Group](#), Iowa State University and [Lushbough Bioinformatics Group](#), University of South Dakota

## Table of Contents

<b>Introduction .....</b>	<b>3</b>
<b>Querying Data Sources .....</b>	<b>4</b>
EXECUTING A QUERY .....	4
<i>Creating a Query</i> .....	4
<i>Executing a Wild Card Search</i> .....	5
VIEWING QUERY RESULTS .....	5
<i>Viewing a Set of Records</i> .....	5
<i>Filtering a Set of Records</i> .....	6
<i>Exporting a Set of Records</i> .....	6
<i>Saving a Set of Records</i> .....	7
<b>BioExtract Server Analytic Tools .....</b>	<b>9</b>
INTRODUCTION .....	9
EXECUTING AN ANALYTIC TOOL .....	10
<i>Creating Data Extracts Using Analytic Tools</i> .....	10
<i>A Second Example With BioMart and Fetch Sequence Records</i> .....	13
ADDING AN ANALYTIC TOOL .....	17
<i>Selecting From a List of Available Tools</i> .....	17
<i>Modifying Added Tools</i> .....	19
<i>Adding a Local Tool</i> .....	21
<i>Running a Local Tool</i> .....	30
<b>BioExtract Server Workflows .....</b>	<b>32</b>
INTRODUCTION .....	32
CREATING A WORKFLOW .....	33
EXECUTING A WORKFLOW .....	42
MODIFYING A WORKFLOW .....	43
COPYING A WORKFLOW .....	46
<b>BioExtract Server Groups.....</b>	<b>47</b>
ABOUT GROUPS .....	47
CREATING A GROUP .....	48
ADDING TOOLS, WORKFLOWS AND DATA EXTRACTS TO A GROUP .....	49
REMOVING TOOLS, WORKFLOWS AND DATA EXTRACTS FROM A GROUP .....	51
INVITING MEMBERS TO JOIN A GROUP .....	52
<b>Workflows through MyExperiment .....</b>	<b>56</b>
ABOUT MYEXPERIMENT .....	56
IMPORTING A BIOEXTRACT SERVER WORKFLOW INTO MYEXPERIMENT .....	56

## INTRODUCTION

The BioExtract Server ([bioextract.org](http://bioextract.org)) is an open, Web-based system designed to aid researchers in the analysis of genomic data by providing a platform for the creation of bioinformatics workflows. Scientific workflows are created within the system by recording tasks performed by the user. These tasks may include querying multiple, distributed data sources, saving query results as searchable data extracts, and executing local and Web-accessible analytic tools. The series of recorded tasks can then be saved as a reproducible, sharable workflow available for subsequent execution with the original or modified inputs and parameter settings. Integrated data resources include interfaces to the National Center for Biotechnology Information (NCBI) nucleotide and protein databases, the European Molecular Biology Laboratory (EMBL-Bank) non-redundant nucleotide database, the Universal Protein Resource (UniProt), and the UniProt Reference Clusters (UniRef) database. The system offers access to numerous preinstalled, curated analytic tools and also provides researchers with the option of selecting computational tools from a large list of Web services including the European Molecular Biology Open Software Suite (EMBOSS), BioMoby, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The system further allows users to integrate local command line tools residing on their own computers through a client-side Java applet.

## QUERYING DATA SOURCES

### Executing a Query

#### Creating a Query

Queries are constructed in the BioExtract Server by first checking the box for one or more data sources and then composing a query using the Query Form. The screen grab below presents an example of querying the NCBI Protein Database for the R2R3-MYB gene in *Pinus taeda* (Loblolly pine) (Figure 1). Queries executed against data sources residing at the BioExtract Server respond quickly. For example, a query of *Viridiplantae* returns in a matter of seconds. Response times for queries against data sources residing at other sites (e.g. NCBI Nucleotide and Protein databases) are consistent with response times for queries made directly at those sites.

**BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ [sign out](#) ]

[Query](#) [Extracts](#) [Tools](#) [Workflows](#) [Groups](#) [Help](#) [Demo Workflow](#)

**Available Data Sources.** Select one or more data sources to query:

- ☐ All
- ☐ Miscellaneous
- ☐ Nucleotide Sequences
- ☒ Protein Sequences
  - ☒ NCBI Protein Database
  - ☐ UniRef
  - ☐ UniProtKB
- ☐ Viridiplantae
- ☐ Viridiplantae Protein

**Fetch Sequence(s)**  
If you know the GI/accession number of the sequence(s) you want to retrieve, use the [Fetch Sequence Records](#) tool. Retrieved records will display on the Extracts page.

**What's New**

**Query Form.** Select a search field and enter a search term.  
Press Add Search Line to combine search terms with AND, OR, AND NOT. [Query examples.](#)

Add Search Line

	Search Field	Search Term(s)
✗	Definition	R2R3-MYB
✗	AND	
	Taxonomy	Pinus taeda


**Current Query:** Definition='R2R3-MYB ' AND Taxonomy='Pinus taeda '

[Submit Query](#) [Clear](#)

Version: 2.3.06 (release)

Figure 1. The BioExtract Server Query page showing the data source NCBI Protein Database selected and the query form set-up to find the R2R3-MYB gene in *Pinus taeda*.

When attempting to locate sequence records associated with a fairly long description, you may find it helpful to use Boolean operators to break down the description into smaller units. As an example, suppose you are interested in locating nucleotide records containing the description *Homo sapiens zinc finger BTB domain*. Using the Boolean operator AND, this could be broken into *Homo sapiens* **AND** *zinc finger* **AND** *BTB domain* as illustrated in the screen grab below (Figure 2).


**BioExtract Server**  
 data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
 Welcome, [clushbou@gmail.com](#) [ [sign out](#) ]



[Query](#) | [Extracts](#) | [Tools](#) | [Workflows](#) | [Groups](#) | [Help](#)

Demo Workflow

**Available Data Sources.** Select one or more data sources to query:
 

☒ All  
☐ Miscellaneous  
☒ Nucleotide Sequences  
☒ Protein Sequences  
☒ Viridiplantae  
☒ Viridiplantae Protein

Nucleotide (nucore)

**Fetch Sequence(s)**  
 If you know the GI/accession number of the sequence(s) you want to retrieve, use the [Fetch Sequence Records](#) tool. Retrieved records will display on the Extracts page.  
**What's New**  
 

**Query Form.** Select a search field and enter a search term.  
 Press Add Search Line to combine search terms with AND, OR, AND NOT. [Query examples](#).

Add Search Line

	Search Field		Search Term(s)
✗	Definition	=	Homo sapiens
✗	AND	Definition	BTB domain
✗	AND	Definition	zinc finger

**Current Query:** Definition='Homo sapiens' AND Definition='BTB domain' AND Definition=' zinc finger '

[Submit Query](#)
[Clear](#)

Figure 2. The Query page showing the Query Form using the Boolean operator AND to find *Homo sapiens zinc finger BTB domain*. Search terms can be combined with Boolean operators to create precise queries.


## Executing a Wild Card Search

The BioExtract Server offers limited wild card searching functionality where the asterisk “\*” is used to represent any character. For example, to locate *Arabidopsis thaliana basic helix-loop-helix (bHLH)* family of proteins, the taxonomy search value would be specified as “*Arabidopsis thaliana*” and the gene value on which to search would be “bHLH\*”. This query will return all *Arabidopsis thaliana* protein records that have a gene annotation beginning with bHLH.

## Viewing Query Results

### Viewing a Set of Records

The results of a query are displayed on the Extracts page. For each record in the result set, there are two links. The Local Details “view record” link provides quick access to details pertaining to the selected record. The External Link allows navigation to the Website hosting the original data record (Figure 3).


**BioExtract Server**  
data access, analysis, storage, and workflow creation


Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

Query Extracts Tools Workflows Groups Help

Demo Workflow

Query returned 13 records.

Databases

Database	Total	Database Description
 Protein	13	Protein sequence record

Select Records Export Records Save Extract

Results Page: << first < prev 1 next > last >>













External Link	Local Details	Description
 <a href="#">223674401</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB22 [Pinus taeda]
 <a href="#">223674399</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB21 [Pinus taeda]
 <a href="#">223674397</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB16 [Pinus taeda]
 <a href="#">223674395</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB13 [Pinus taeda]
 <a href="#">223674393</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB10 [Pinus taeda]
 <a href="#">223674391</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB5 [Pinus taeda]
 <a href="#">34147926</a>	<a href="#">view record</a>	R2R3-MYB transcription factor [Pinus taeda]
 <a href="#">34147924</a>	<a href="#">view record</a>	R2R3-MYB transcription factor [Pinus taeda]
 <a href="#">89058608</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB2 [Pinus taeda]
 <a href="#">89058606</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB3 [Pinus taeda]
 <a href="#">89058604</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB7 [Pinus taeda]
 <a href="#">89058602</a>	<a href="#">view record</a>	R2R3-MYB transcription factor MYB8 [Pinus taeda]

Figure 3. The BioExtract Server Extracts page showing the results of a query performed on the Query page. Clicking a Local Details link displays the selected record's GenBank file. Clicking an External Link takes you to the Website hosting the original data record.

### Filtering a Set of Records

Records under the Extracts tab may be easily filtered to specific subsets by using the *Select Records* button. Clicking the check box to the left of a desired record adds that record to the subset. Clicking the *Select All on Page* button selects all of the records on an individual page. Clicking the *Complete* button after selecting all desired records generates a new result set containing only the filtered portion.

### Exporting a Set of Records

Records under the Extracts tab may be downloaded using the *Export Records* button. The *Export Records* feature has options for downloading records in FASTA format or in the default format, which depends on the data source queried (Figure 4). For example, if an NCBI data source were queried, the default format would be GenBank.

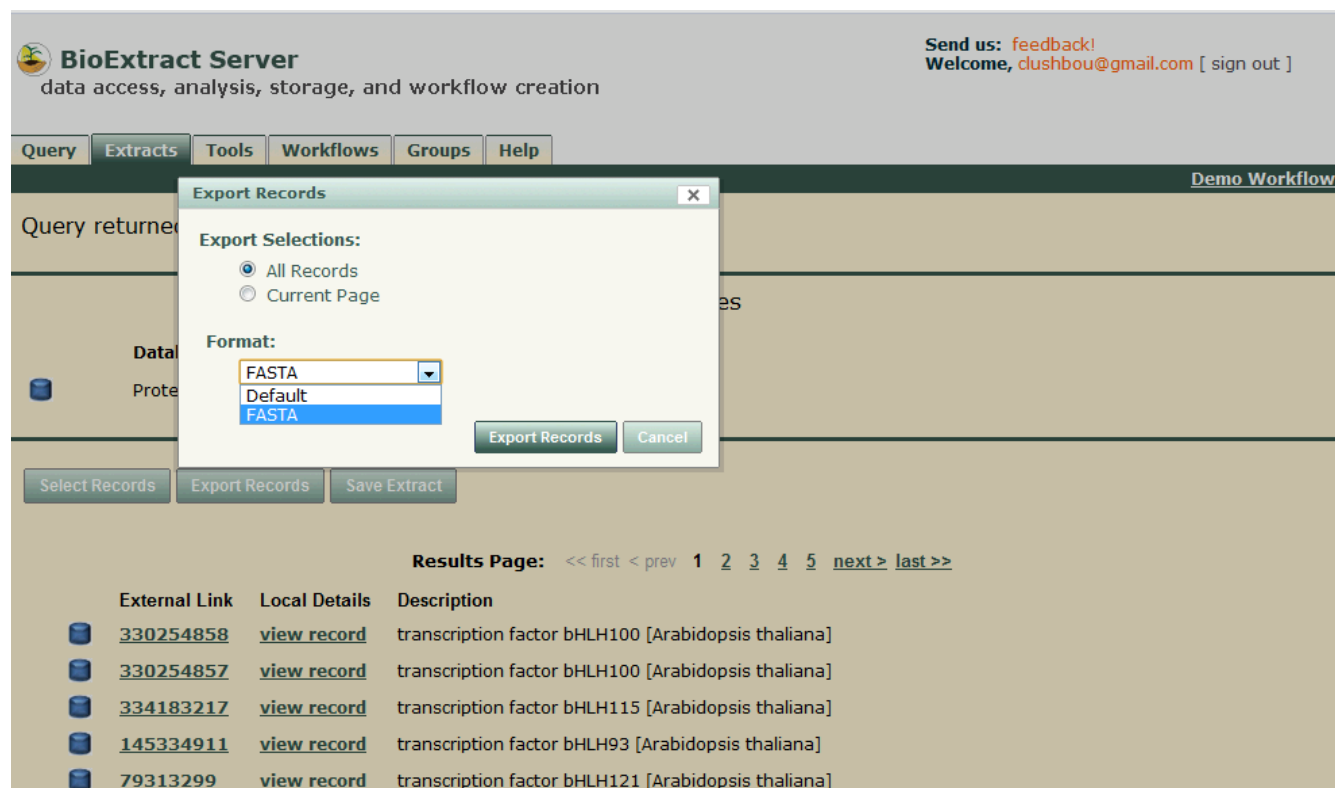


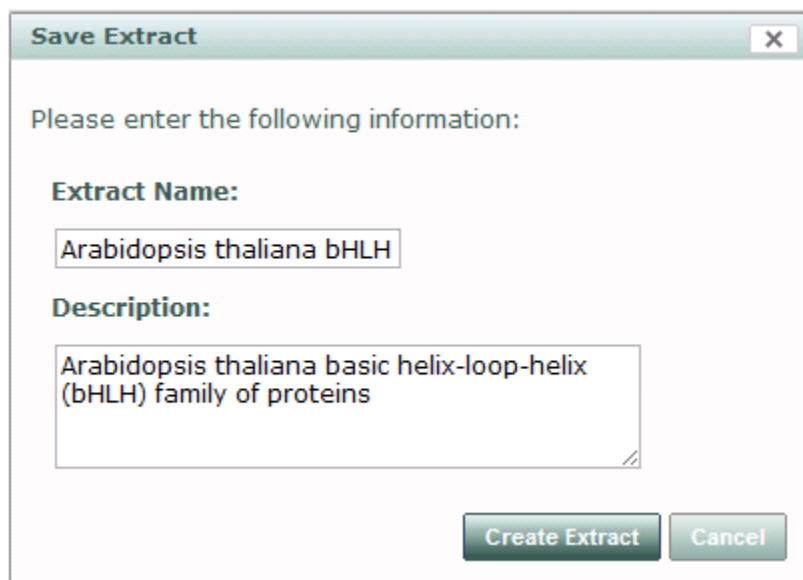
Figure 4. The BioExtract Server Extracts page showing the Export Records feature with downloading options FASTA and Default, which depends on the data source queried.

### Saving a Set of Records

While most of the BioExtract Server’s functionality is available to all users, the ability to save records on the BioExtract Server website is available only to users who have registered with the system.

Users who are signed into the BioExtract Server may save records as searchable data extracts by using the *Save Extract* button available from the Extracts page (Figure 5). Once saved, data extracts can be found on the Query page in Available Data Sources under the heading *Miscellaneous* (Figure 6).

All data extracts saved on the BioExtract Server website are privately owned by the user and are only made available to others by explicitly sharing them with a group. This is accomplished by: 1) clicking the Groups tab, 2) creating a group under additional actions, 3) clicking the new group name, 4) selecting the Extracts tab for the new group, and finally 5) clicking the “Add Elements” button to select the data extract to share.



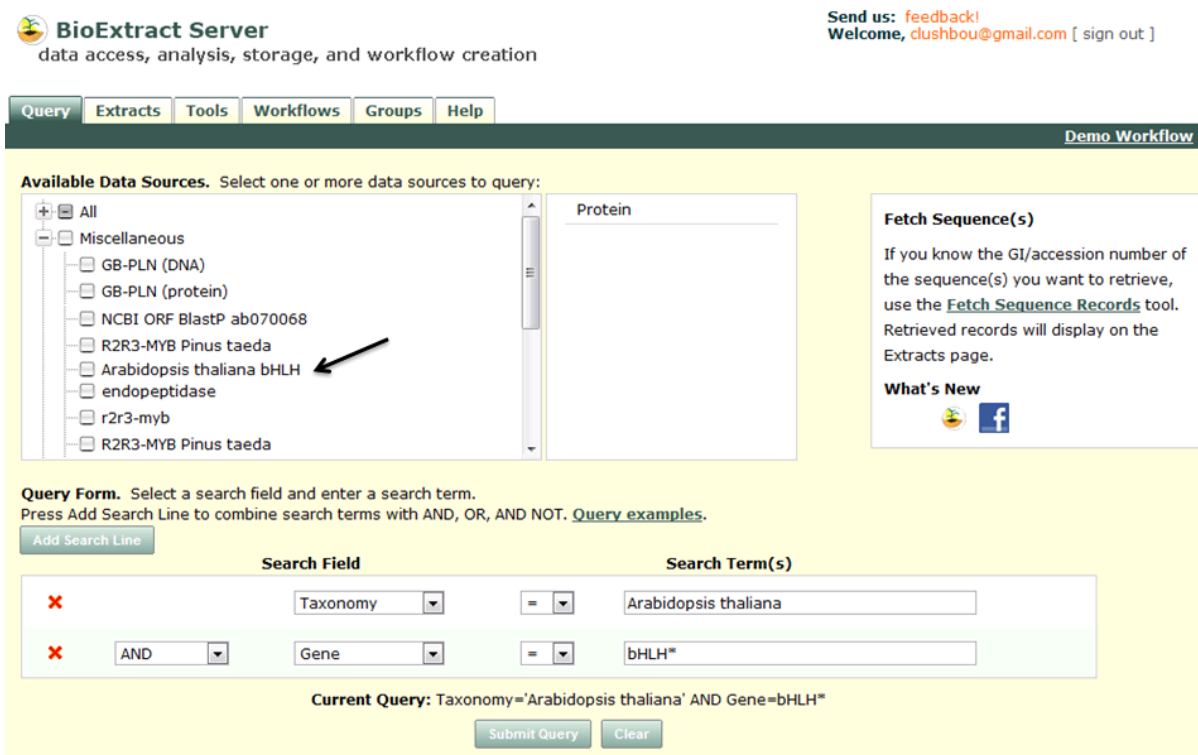
**Save Extract** [X]

Please enter the following information:

**Extract Name:**

**Description:**

Figure 5. The Save Extract window showing Extract Name and Description entered. Clicking the *Create Extract* button creates a data extract which is stored on the BioExtract Server website.



**BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]



Query Extracts Tools Workflows Groups Help Demo Workflow

**Available Data Sources.** Select one or more data sources to query:

- ☒ All
- ☐ Miscellaneous
  - ☐ GB-PLN (DNA)
  - ☐ GB-PLN (protein)
  - ☐ NCBI ORF BlastP ab070068
  - ☐ R2R3-MYB Pinus taeda
  - ☒ Arabidopsis thaliana bHLH
  - ☐ endopeptidase
  - ☐ r2r3-myb
  - ☐ R2R3-MYB Pinus taeda

Protein

**Fetch Sequence(s)**  
If you know the GI/accession number of the sequence(s) you want to retrieve, use the [Fetch Sequence Records](#) tool. Retrieved records will display on the Extracts page.

**What's New**  
 

**Query Form.** Select a search field and enter a search term.  
Press Add Search Line to combine search terms with AND, OR, AND NOT. [Query examples.](#)

	Search Field		Search Term(s)
×	Taxonomy	=	Arabidopsis thaliana
×	AND	Gene	bHLH*

**Current Query:** Taxonomy='Arabidopsis thaliana' AND Gene=bHLH\*

Figure 6. The Query page showing a user's privately owned data extract *Arabidopsis thaliana bHLH* in Available Data Sources under the heading Miscellaneous.



## BIOEXTRACT SERVER ANALYTIC TOOLS

### Introduction

A number of well-established and unique bioinformatics analytic tools are made available through the BioExtract Server, with the majority integrated as curated Web services. Users access analytic tools through the list of Available Tools on the Tools page (Figure 7). The source of the analytic tool's input may be: 1) the records listed on the Extracts page, 2) the output from a previously executed tool, or 3) private data provided by the user (uploaded or entered in a text box). Analytic tool parameters may be selected or modified before execution and resulting output files may be viewed, downloaded and used as input into subsequently executed analytic tools.

The BioExtract Server offers users the ability to add analytic tools to their BioExtract Server workspace. Users may select such tools from a list of web services, including EMBOSS SoapLab, BioMoby and KEGG, with the integration of BioMart (www.biomart.org) currently in development. The BioExtract Server also offers generic support for other SOAP-based Web services and lets users integrate local command line tools residing on their own workstations through the use of a client-side Java applet. Analytic tools added by users may be annotated through the "Customize Your Tools" functionality, which allows users to provide detailed descriptions of the tools, as well as add a help link to additional information.

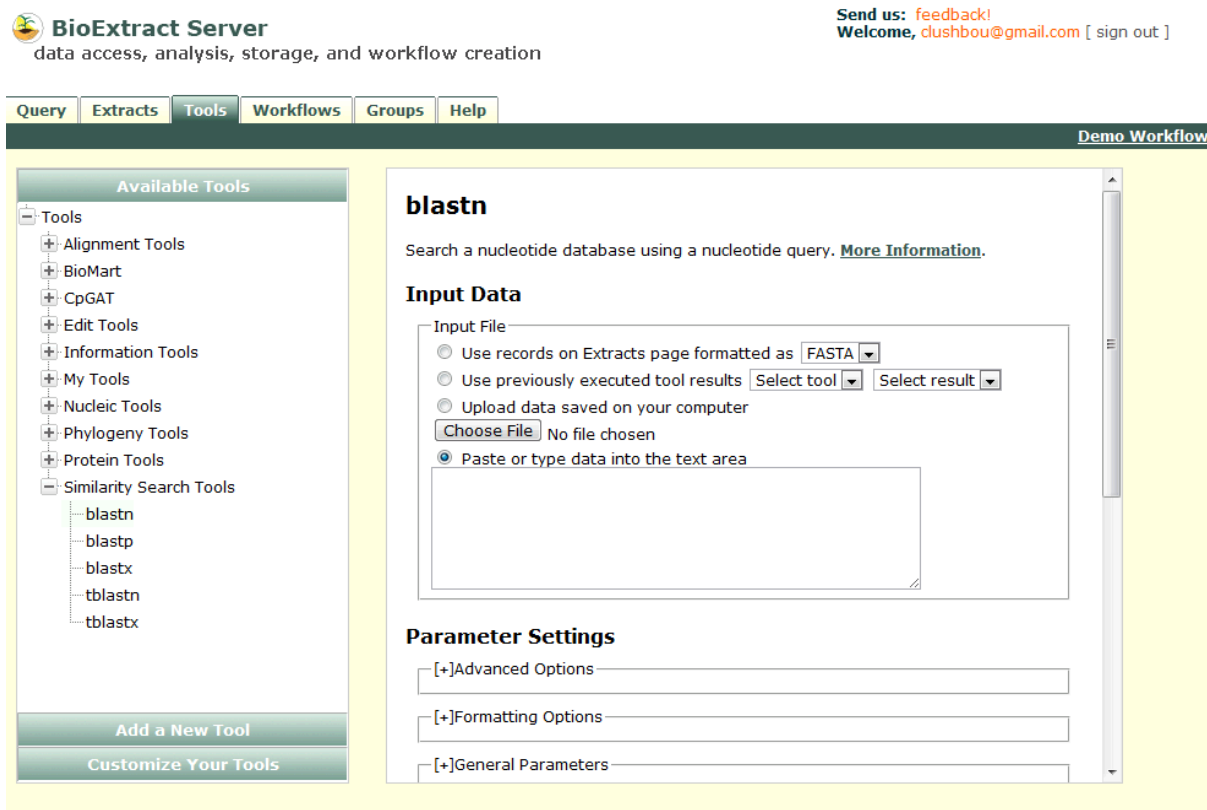


Figure 7. The BioExtract Server Tools Page showing the Available Tools with the Similarity Search Tools group expanded and the blastn tool selected.

## Executing an Analytic Tool


Analytic tools are executed in the BioExtract Server by first selecting a tool from the list of Available Tools on the Tools page. One of the most important characteristics of the BioExtract Server is allowing the user to specify the input source when executing a tool. For each tool, four possible input methods are available: using the records listed on the Extracts page, using the output from a previously executed tool, entering data into a text box and uploading a data file. When executing a tool using the “Use records listed on the Extracts page” option, the default file format is FASTA. For each record listed, the BioExtract Server system will retrieve it in FASTA format and create an input file for the tool.

Attempts have been made in the BioExtract Server to provide an appropriate level of abstraction to hide as much low-level format transformation as possible. In situations where this is not possible, users may run intermediate tools or “shims” to perform the necessary format conversions. The BioExtract Server has incorporated a predefined tool for format conversion and many of the Web services available to users have been explicitly defined for data filtering and transformation.


As an example, if the output from one analytic tool is in FASTA format and the input into another tool is required to be in GenBank format, the user may run the intermediate tool or “shim”, *FormatConversion* to convert the FASTA formatted file to GenBank format. *FormatConversion* makes the conversion by parsing the id from the FASTA file and retrieving the record in GenBank format.

## Creating Data Extracts Using Analytic Tools

The execution of some analytic tools results in the creation of a data extract, which displays on the Extracts page. As an example, let’s create a data extract using a *Gallus gallus cadherin-19* protein sequence record and the analytic tool BLAST [25]. First, we’ll query the NCBI Protein Database for Species = *Gallus gallus* and Definition = *cadherin 19*. From the query results, use the Select Records button to specify one record as the sequence of interest.

	External Link	Local Details	Description
	<a href="#">154152095</a>	<a href="#">view record</a>	cadherin-19 [Gallus gallus]

Next, click the Tools tab, open the “Similarity Search Tools” group and select the *blastp* tool. The tool form opens in the right panel. In the Input Data section select “Use records on Extracts page formatted as FASTA” and click the *Execute* button.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Current User: [guest](#)  
[ [sign in](#) | [register](#) | [why register?](#) ]

Query | Extracts | **Tools** | Workflows | Groups | Help

Demo Workflow

Available Tools

Tools

+

 Alignment Tools

+

 BioMart

+

 CpGAT

+

 Edit Tools

+

 Information Tools

+

 Nucleic Tools

+

 Phylogeny Tools

+

 Protein Tools

-

 Similarity Search Tools

- blastn
- blastp
- blastx
- tblastn
- tblastx

Add a New Tool

Customize Your Tools

**blastp**

Search a protein database using a protein query. [More Information.](#)

**Input Data**

Input File

☒ Use records on Extracts page formatted as FASTA

☐ Use previously executed tool results 1 blastp Select result

☐ Upload data saved on your computer  
Choose File No file chosen

☐ Paste or type data into the text area

**Parameter Settings**

[ - ]Filters and Masking


Filter: Low-complexity ( FILTER L ) ☐

Filter: Human repeats ( FILTER R ) ☐

Filter: Mask for lookup table only ( FILTER m ) ☐

## Version 1.0

A data extract will be created on the Extracts page containing sequence records similar to the *Gallus gallus cadherin-19* sequence. This data extract can be saved on the BioExtract Server website, shared with others and used as input into other analytic tools.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation


Send us: [feedback!](#)  
Current User: [guest](#)  
[ [sign in](#) | [register](#) | [why register?](#) ]

Query | **Extracts** | Tools | Workflows | Groups | Help

[Demo Workflow](#)














Query returned **50** records.

Databases

Database	Total	Database Description
 NCBI	50	

Select Records | Export Records | Save Extract

Results Page: << first < prev 1 next > last >>

External Link	Local Details	Description
 <a href="#">NP_001093757</a>	<a href="#">view record</a>	cadherin-19 [gallus gallus]
 <a href="#">XP_003204982</a>	<a href="#">view record</a>	predicted: cadherin-19-like [meleagris gallopavo]
 <a href="#">XP_002191433</a>	<a href="#">view record</a>	predicted: cadherin 19, type 2 [taeniopygia guttata]
 <a href="#">XP_001372408</a>	<a href="#">view record</a>	predicted: cadherin-19 [monodelphis domestica]
 <a href="#">XP_001094512</a>	<a href="#">view record</a>	predicted: cadherin-19 isoform 2 [macaca mulatta]
 <a href="#">XP_523961</a>	<a href="#">view record</a>	predicted: cadherin-19 isoform 2 [pan troglodytes]
 <a href="#">XP_002757363</a>	<a href="#">view record</a>	predicted: cadherin-19 [callithrix jacchus]
 <a href="#">XP_002922100</a>	<a href="#">view record</a>	predicted: cadherin-19-like [ailuropoda melanoleuca]
 <a href="#">XP_001491702</a>	<a href="#">view record</a>	predicted: cadherin 19, type 2 [equus caballus]
 <a href="#">NP_066976</a>	<a href="#">view record</a>	cadherin-19 preproprotein [homo sapiens]
 <a href="#">XP_541068</a>	<a href="#">view record</a>	predicted: similar to cadherin 19, type 2 preproprotein [canis]
 <a href="#">XP_003264372</a>	<a href="#">view record</a>	predicted: cadherin-19 [nomascus leucogenys]
 <a href="#">XP_003219759</a>	<a href="#">view record</a>	predicted: cadherin-7-like [anolis carolinensis]


## A Second Example With BioMart and Fetch Sequence Records

Now let's create a new data extract using BioMart and Fetch Sequence Records tools.


1. From the Tools page, expand the BioMart group and select ENSEMBL GENES 65 (SANGER UK).
2. From the expanded group, select the *Homo sapiens genes (GRCh37.p5)* dataset. The tool form opens.

The screenshot displays the BioExtract Server interface. At the top, the header includes the BioExtract Server logo, the tagline 'data access, analysis, storage, and workflow creation', and user information: 'Send us: feedback!', 'Current User: guest', and links for '[ sign in | register | why register? ]'. Below the header is a navigation bar with tabs for 'Query', 'Extracts', 'Tools', 'Workflows', 'Groups', and 'Help'. A 'Demo Workflow' link is also present on the right. The main content area is divided into two panels. The left panel, titled 'Available Tools', contains a scrollable list of datasets. An arrow points to the entry 'Homo sapiens genes (GRCh37.p3)'. The right panel, titled 'Homo sapiens genes (GRCh37.p3)', shows the tool configuration form. It includes buttons for 'Count', 'Execute', and 'View Results'. Below these are two tabs: 'Filters' and 'Attributes'. The 'Attributes' tab is active, showing a list of attributes with expandable options (indicated by '+' icons): REGION, GENE, TRANSCRIPT EVENT, GENE ONTOLOGY, EXPRESSION, MULTI SPECIES COMPARISONS, PROTEIN DOMAINS, and VARIATION.

3. Narrow the results by clicking the Region filter, specifying the X chromosome and q28 Band Start and Band End.

REGION: 

☒ Chromosome



☐ Base pair

Gene Start (bp)

Gene End (bp)

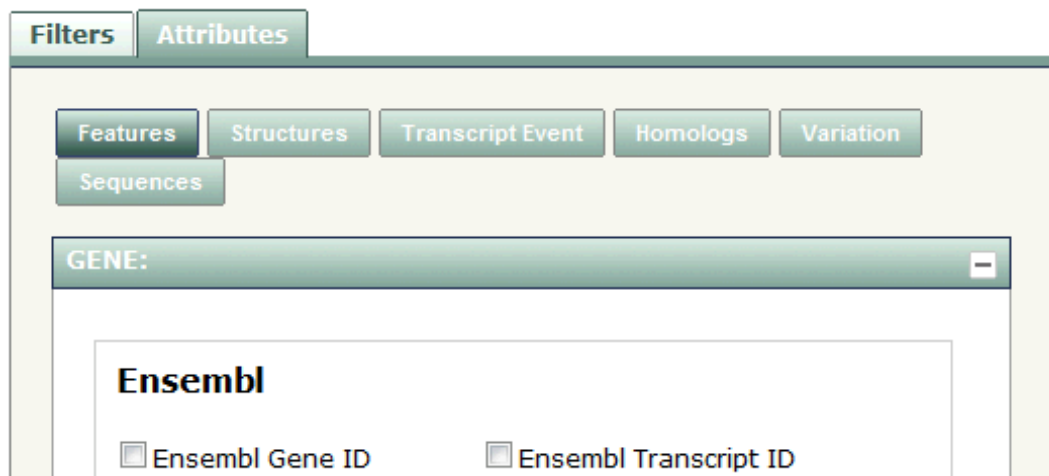
☒ Band

Band Start

Band End

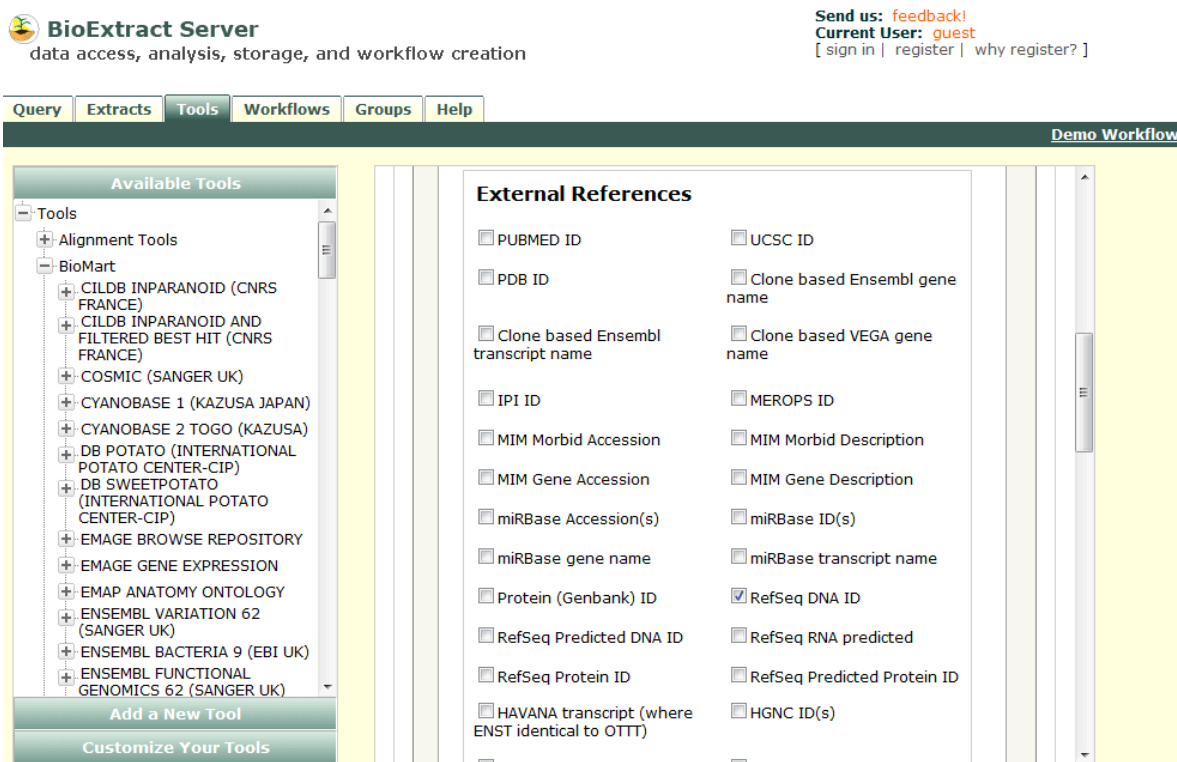
4. Click the **Attributes** tab and select the **Gene** filter.

- a. In Ensembl, deselect *Ensembl Gene ID* and *Ensembl Transcript ID*.



5. Click the **External** filter.

- b. In External References, select *RefSeq DNA ID*.



6. Scroll to the top of the tool form and click the *Execute* button.



7. Now open and run the Fetch Sequence Records tool to create a data extract. From the Tools page, expand the Information Tools group and select the *Fetch Sequence Records* tool. The tool form will open in the right panel.

8. In the Input Data section, select “Use previously executed tool results”. In the associated drop-down menu, select “BioMart Query”.

9. In Parameter Settings > Database Options, set the database to *refseq*, then click the *Execute* button. A data extract will be created on the Extracts page. This data extract can be saved on the BioExtract Server website, shared with others and used as input into other analytic tools.

**BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Current User: [guest](#)  
[ [sign in](#) | [register](#) | [why register?](#) ]

Query Extracts Tools Workflows Groups Help

[Demo Workflow](#)

Query returned **249** records.

Databases		
Database	Total	Database Description
NCBI	249	

Select Records Export Records Save Extract

Results Page: << first < prev 1 2 3 4 5 next > last >>

External Link	Local Details	Description
<a href="#">NR_027456</a>	<a href="#">view record</a>	homo sapiens hypothetical loc100272228 (loc100272228), non-coding rna
<a href="#">NM_021049</a>	<a href="#">view record</a>	homo sapiens melanoma antigen family a, 5 (magea5), mrna
<a href="#">NM_001129826</a>	<a href="#">view record</a>	homo sapiens csag family, member 3 (csag3), transcript variant 1, mrna
<a href="#">NM_001080848</a>	<a href="#">view record</a>	homo sapiens csag family, member 2 (csag2), transcript variant 1, mrna
<a href="#">NM_001129826</a>	<a href="#">view record</a>	homo sapiens csag family, member 3 (csag3), transcript variant 1, mrna
<a href="#">NM_001080848</a>	<a href="#">view record</a>	homo sapiens csag family, member 2 (csag2), transcript variant 1, mrna
<a href="#">NM_001136273</a>	<a href="#">view record</a>	homo sapiens zinc finger protein 92 homolog (mouse) (zfp92), mrna
<a href="#">NM_020061</a>	<a href="#">view record</a>	homo sapiens opsin 1 (cone pigments), long-wave-sensitive (opn1lw), mrna
<a href="#">NM_000513</a>	<a href="#">view record</a>	homo sapiens opsin 1 (cone pigments), medium-wave-sensitive (opn1mw), mrna
<a href="#">NM_001048181</a>	<a href="#">view record</a>	homo sapiens opsin 1 (cone pigments), medium-wave-sensitive 2 (opn1mw2), mrna
<a href="#">NM_001586</a>	<a href="#">view record</a>	homo sapiens testis expressed 28 (tex28), transcript variant 2, mrna
<a href="#">NM_005342</a>	<a href="#">view record</a>	homo sapiens high mobility group box 3 (hmgbox3), mrna



## Adding an Analytic Tool

### Selecting From a List of Available Tools

Adding an analytic tool by selecting from a list of Web services is fairly simple, primarily because most of the information required by the BioExtract Server system can be obtained by parsing the XML Web service description and translating it into a BioExtract Tool Object. The particular GUI control that the system maps to a specific tool parameter is a function of the parameter type.

From the Tools page, use the “Add a New Tool” menu to view the list of tools, which you can add to your BioExtract Server account. The “search” text box above the list of tools lets you search for a tool by name. As an example, entering *needle* will highlight those tools containing the word “needle” (Figure 8). Clicking on the *Save Tool* button adds the tool to your BioExtract Server account.

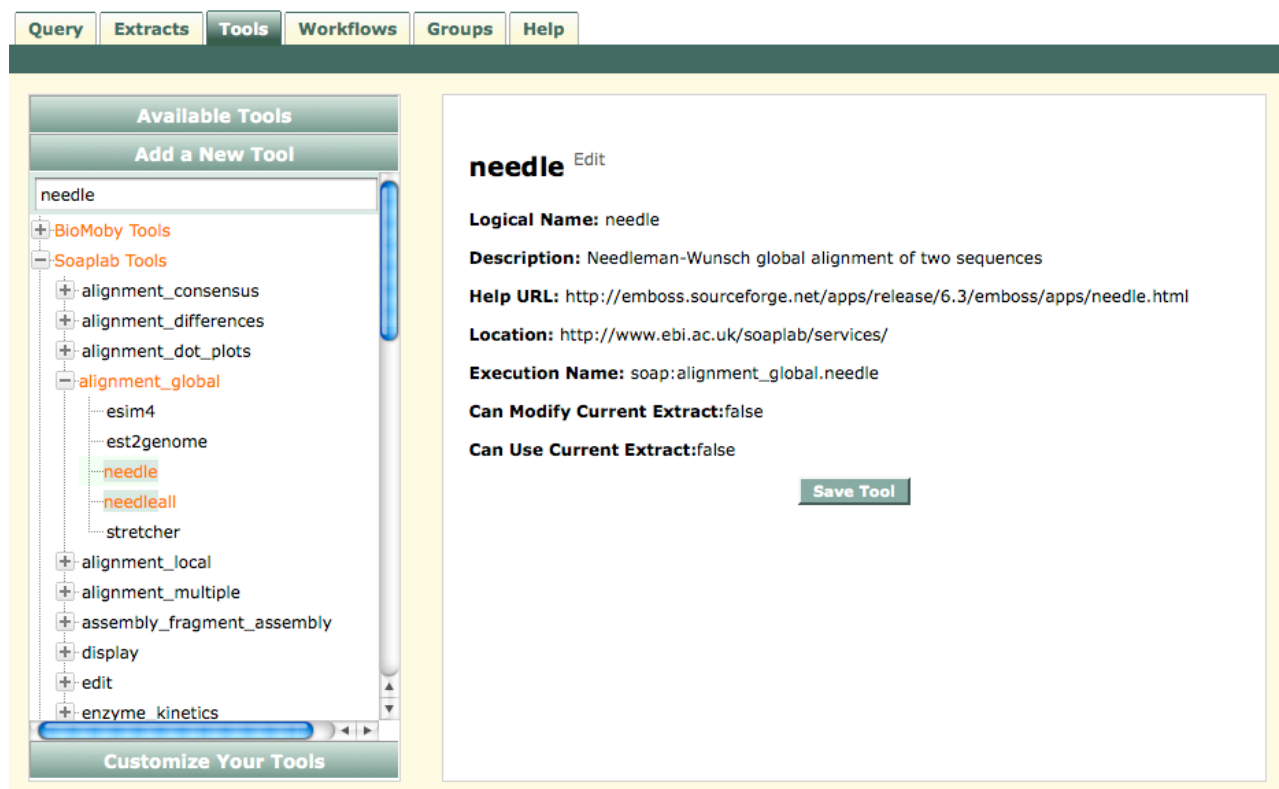
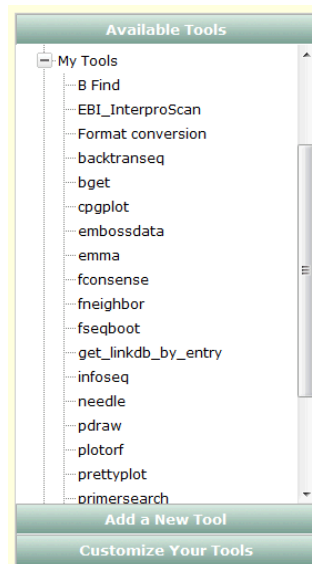


Figure 8. The BioExtract Server Tools page showing the “Add a New Tool” menu expanded with a search for tools containing the word *needle*. In this example, the Soaplab *needle* tool has been selected.

## Version 1.0

Added tools can be found on the Tools page in the “Available Tools” menu in the “My Tools” group.

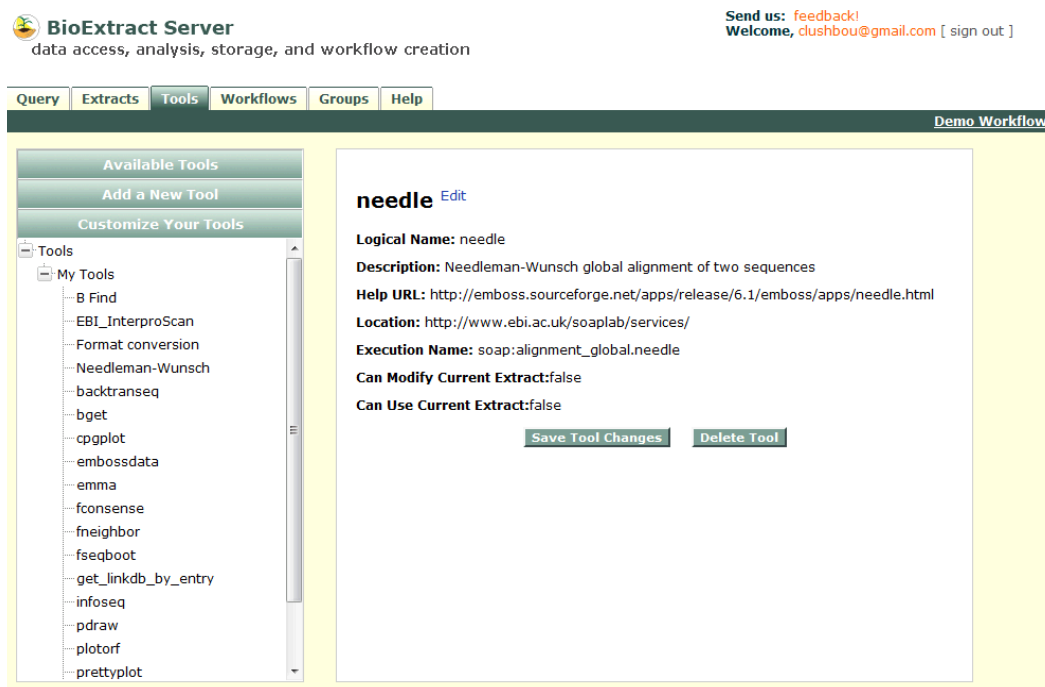


## Modifying Added Tools

Users have the option to change the logical names of the tool attributes (e.g., parameters, inputs and tool name), which allows them to customize the tool interface. If the input to the tool being added represents sequence data, the user may indicate that the tool can use the BioExtract Server result set as input.

### To modify an added tool:

1. From the Tools page, click the “Customize Your Tools” menu and select the tool to be modified. The selected tool opens in the right panel.



2. Click the **Edit** link adjacent to the tool name. The tool form opens.

## Version 1.0

In the tool form, you can change the Logical Name, Description and HelpURL. As an example, let's modify the tool name by entering *Needleman-Wunsch* in the Logical Name box.

The screenshot displays the BioExtract Server web interface. At the top, the header includes the BioExtract Server logo and tagline 'data access, analysis, storage, and workflow creation', a user welcome message 'Welcome, clushbou@gmail.com', and a 'sign out' link. A navigation bar contains tabs for 'Query', 'Extracts', 'Tools', 'Workflows', 'Groups', and 'Help'. The 'Tools' tab is active, showing a 'Demo Workflow' button. On the left, a sidebar lists 'Available Tools' with options to 'Add a New Tool' or 'Customize Your Tools'. Under 'My Tools', a list of tools is shown, including 'needle'. The main panel displays the 'needle' tool configuration form. It includes a 'Save' link and a 'Cancel' button. The 'Logical Name' field contains 'needle'. The 'Description' field contains 'Needleman-Wunsch global alignment of two sequences'. The 'HelpURL' field contains 'http://emboss.sourceforge.net/apps/release/6.1/emboss/apps/r'. The 'Can Use Current Extract' checkbox is unchecked. At the bottom of the form are 'Save Tool Changes' and 'Delete Tool' buttons.

3. Click the **Save** link at the top of the tool form to keep your changes. Then click the **Save Tool Changes** button to complete the process.

## Adding a Local Tool

A local tool is a program on your own computer. The tool itself is not uploaded to the BioExtract Server. Rather, the BioExtract Server uses the information given about the tool to execute it on your own system. Once the tool has been added, it can be used like any of the other tools in the BioExtract Server and can even be included in workflows.

The tool must meet the following criteria:

- The tool can be executed from the command line (either a DOS prompt or Linux shell). If the tool only has a graphical or window-like interface, BioExtract Server cannot use it.
- If you wish to use the output of the local tool as input for another tool on the BioExtract Server, the name and location of any output files produced by the tool must be known before the tool is executed.

If the tool meets the criteria above, the following information is required to add the tool:

- The full path to the location of the tool on your system. For example:  
"C:\biotools\mytool.exe" on Windows  
"/usr/local/bio/mytool" on Linux
- Any required command line arguments. These are pieces of information typed after the program's name, usually specifying the name of the file it should use as input, the file name its output should be written to, etc. For example, let's say we have a tool that can convert sequences in one format to another format. The documentation for the tool may read: usage: seqconv -i filename -if formatName -of formatName  
Where:  
-i filename is the file to be converted  
-if is the format of the input file  
-of is the requested format of the output file  
Valid format names: fasta, blast, genbank We have a file in fasta format, located at "C:\temp\fasta.txt", that we wish to convert to the genbank format. According to the documentation, we could use the following command to accomplish this: seqconv -i C:\temp\fasta.txt -if fasta -of genbank
- Any additional requirements or constraints the tool may have in regards to input and output files. For example, some tools may require that all files used for input be located in a particular directory, while others may allow files in any location. Some tools may allow you to specify the name and location for each output file, while others may automatically give a name to the file and place it in a predetermined location.

**As an example**, suppose we have a tool called "Fetch Records", which takes a file containing a list of ids and the name of a data source (i.e. ncbi, embl, refseq). The tool accesses the data source and retrieves the records associated with the ids.

The screenshot shows the BioExtract Server web interface. At the top, there's a header with the BioExtract Server logo and tagline 'data access, analysis, storage, and workflow creation'. On the right, there's a user login area with 'Send us: feedback!', 'Welcome, clushbou@gmail.com', and a 'sign out' link. Below the header is a navigation bar with tabs: Query, Extracts, Tools, Workflows, Groups, and Help. The 'Tools' tab is selected. On the right side of the navigation bar, there's a 'Demo Workflow' link. The main content area is divided into two panels. The left panel, titled 'Available Tools', contains a search bar and a list of tool categories: BioMoby Tools, Soaplab Tools, KEGG Tools, Add a Local Tool, New Local Tool, and Add a Web Service. The right panel, titled 'New Tool', contains a form for creating a new tool. The form has fields for 'Logical Name', 'Description', 'Help URL', 'Location', 'Execution Name', 'Can Modify Current Extract', and 'Can Use Current Extract'. There are also sections for 'Inputs', 'Outputs', and 'Parameter Groupings', each with a 'Create New' link. A 'Save Tool' button is at the bottom right of the form.

**Step 1:** Click on the Tools tab, then select **Add a New Tool** from the menu that appears on the left.

**Step 2:** Click on **Add a Local Tool**, then **New Local Tool**. Now, click on the **Edit** link next to "New Tool" in the right hand panel.

- **Logical Name:** a name for the tool that will be used in the BioExtract Server. It does not have to match the actual name of the program. For our tool, names such as *Fetch Records*.
- **Description:** Optional. A description of the tool.
- **HelpURL:** Optional. If there is a website associated with your tool, you can enter a link to it here.
- **Location:** Not used for local tools. The field should be left blank.
- **Execution Name:** Enter the full path of your tool here.  
If your tool is a Java JAR file (e.g. ResultSetDB.jar), this should be "java -jar" followed by the full path of your JAR file. For example, on a Linux system the Execution Name might be "java -jar /usr/local/tools/ResultSetDB.jar", and on a Windows system it might be "java -jar C:\ResultSetDB\ResultSetDB\dist\ResultSetDB.jar". Depending on your system configuration, you may need to replace 'java' with the full path to the java command (e.g. C:\Program Files\Java\jre6\bin\java.exe).

- **Can Use Current Extract:** Not used for local tools, and should be left unchecked. As a note, local tools *can* use the current extract as input, but this checkbox has no effect on that at all.

**Step 3:** Once this information is entered, click the **Save** link at the top of the form.

The screenshot shows the BioExtract Server interface. At the top, the logo 'BioExtract Server' is followed by the tagline 'data access, analysis, storage, and workflow creation'. To the right, there is a user greeting: 'Send us: [feedback!](#) Welcome, [clushbou@gmail.com](#) [ sign out ]'. Below this is a navigation bar with tabs: 'Query', 'Extracts', 'Tools' (which is active), 'Workflows', 'Groups', and 'Help'. On the right side of the navigation bar, there is a link for 'Demo Workflow'. The main content area is divided into two panels. The left panel, titled 'Available Tools', contains a search bar and a list of tool categories: 'BioMoby Tools', 'Soaplab Tools', 'KEGG Tools', 'Add a Local Tool' (with a sub-link 'New Local Tool'), and 'Add a Web Service'. At the bottom of this panel is a button labeled 'Customize Your Tools'. The right panel, titled 'New Tool', contains instructions: 'You can change the Logical Name, Description, and HelpURL. After making your changes, click Save at the top of the form, and then click Save Tool Changes to complete the process.' Below the instructions are several input fields: 'Logical Name' (containing 'Fetch Records'), 'Description' (containing 'Given a list of database ids and the name of a database, will retrieve records associated with the ids from the specified database'), 'HelpURL', 'Location', and 'Execution Name' (containing 'java -jar c:\ResultSetDB\ResultSetDB\dist\ResultSetDB.jar'). At the bottom of the form is a checkbox labeled 'Can Use Current Extract:' which is currently unchecked. Above the input fields are 'Save' and 'Cancel' buttons.

**Step 4:** Specify input files.

Data from the BioExtract Server is streamed to the local tool by using one or more input files, created on your computer when BioExtract Server runs the tool. The data stored in each input file can come from one of four sources, selected when the tool is executed:

1. The current extract (records on Extracts page)
2. A previous tool's output
3. An input file uploaded from your computer
4. Text typed or pasted directly into a text box

**Input Data**

input1

☐ Use records on Extracts page formatted as FASTA

☐ Use previously executed tool results Select tool Select result

☐ Upload data saved on your computer Browse...

☒ Paste or type data into the text area

Not all tools require input files, so you may not need to add one at all. Our example tool “Fetch Records” requires a single input file containing a list of Ids representing sequence records in an external data source.

- **Physical Name:** We're going to skip this one for now. It will make more sense later after going through the rest of the fields.
- **Logical Name:** As with the tool itself, the logical name is only used within BioExtract Server. It may help to use the type of data as the logical name (especially if the tool has more than one input). Since Fetch Records expects this input file to be a list of Ids, we'll use *ids* for the logical name.
- **Description:** Any additional information or notes about the tool can be written here.
- **Record Number Limit:** If the current extract is being used as the input, the number of records included will be truncated to this amount. 0 means "no limit".
- **File Name:** Enter the name (with full path) that BioExtract Server should use to create this file. For Fetch Records' input file, we will use *C:\biotools\id\_input.txt*.
- **Data Types:** Not used for local tools.



- **Uses Current Extract:** This control isn't used for local tools, and has no effect at all on whether the input can come from the current extract. It's best to leave it unchecked.

**Inputs** [Create New](#)

---

**New Input**

[Save](#) [Delete](#) [Cancel](#)

**Physical Name:**

**Logical Name:**

**Description:**

**Record Number Limit:**

**File Name:**

**Data Types:**

**Uses Current Extract:** ☐

**File Size Limit:**

**Include in Command Line:** ☒

- **File Size Limit:** If the size of the input file should be restricted to a certain number of bytes, you can enter that number here. This should only be needed if your tool has a limit on the size of the input files given to it. Most tools can accept files of any size, so if you're not sure, it's probably safe to leave this at the default value of 0 (no limit).
- **Include in Command Line:** If your tool expects the name of the input file to be given after its name in the command line, this box should be checked. If **Include in Command Line:** ☐ is checked, BioExtract Server adds the file name to the command used to run to the tool. If **Include in Command Line:** ☐ is left unchecked, the file will still be created with the name

given by **File Name:**  , but its name will not be added to the command line.

- Some tools require a "switch" (like -i or -f) before files given in the command line. For example, Fetch Sequence requires a "-i" before the input file. The command line would contain: “ -i C:\biotools\id\_input.txt “
- Remember the **Physical Name** field we skipped earlier? If the tool requires a switch before the input file, enter the switch in the Physical Name field.
- Once all the information about this input has been entered, click the **Save** link at the top of the form. Any number of inputs can be added, depending on the requirements or limitations of your tool.

#### **Step 5:** Specify output files.

The interface for defining output files is very similar to the one used to define input files. As with input files, output files are optional and need only be defined if you would like to use the output of the tool as input to other tools on BioExtract Server.

Physical Name, Logical Name, Description, File Name, and Include in Command Line behave exactly the same as they do for Input files.

Record Number Limit, Description File Name, and Modify Current Extract are not used for local tool output files and can be ignored.

#### **Step 6:** Command Line Parameters/Arguments

Most tools have a set of options whose values are given by arguments on the command line. Specifying the input and output files for the program is just one example of such arguments. For our purposes here, the terms "argument" and "parameter" are identical.

BioExtract Server uses the information given about each parameter to add an element in the completed tool's interface where the value for the parameter can be given. This is the interface shown for Fetch Records' parameters:

**Parameter Settings for mutate**

Severity ( -sev )

Point Mutations ( -point )

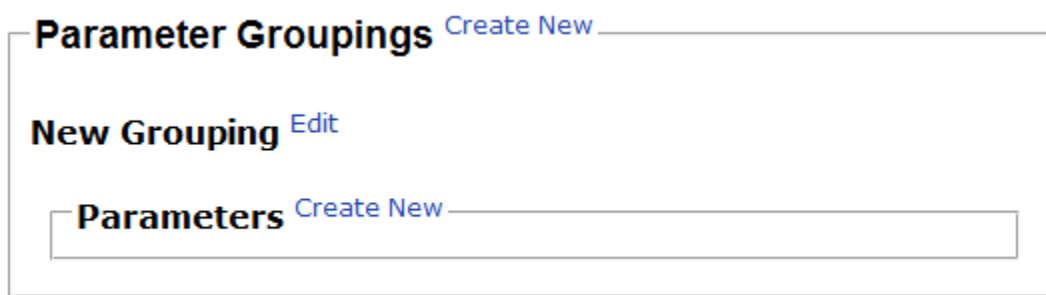
Codon Mutations ( -codon )

Before adding a Parameter, you must first add a Parameter Group. Multiple parameter groups are allowed, and can be used to keep related parameters together. For example, a tool may have a set of parameters

that affect the appearance of the tool's output. These parameters could be placed within a group called "Output Options", and will be displayed together in the menu used to run the tool from BioExtract Server.

Our example tool, Fetch Records, has one parameter, which specifies the name of the database containing the desired sequence records. On the command line it would appear as `-db ncbi`.

- To add parameters, begin by clicking on the **Create New** link next to "Parameter Groupings". Then, click **Edit** next to "New Grouping".

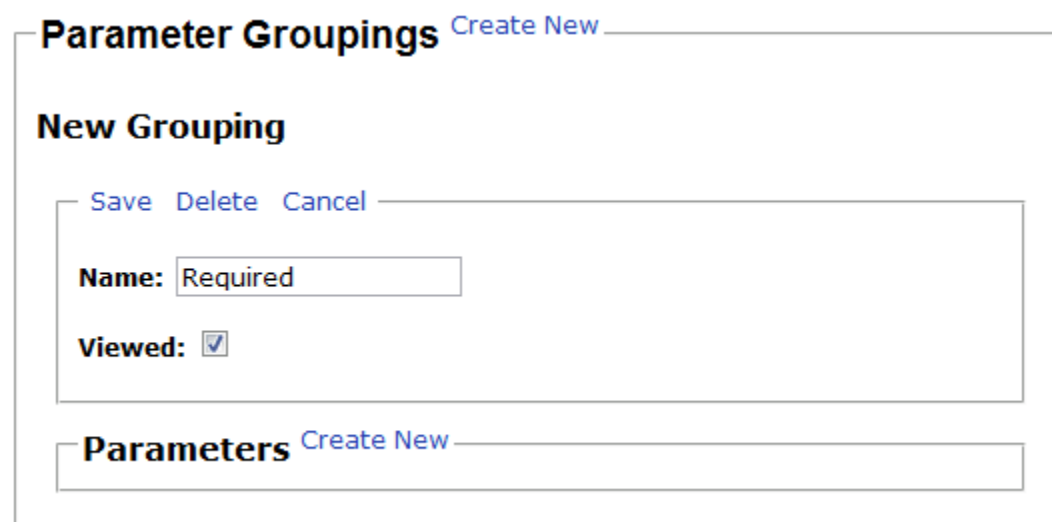


Parameter Groupings [Create New](#)

New Grouping [Edit](#)

Parameters [Create New](#)

- Assign a name to this parameter grouping. The **Viewed:** ☒ check box allows you to specify if you would like to have the parameter group expanded when the tool is being executed.



Parameter Groupings [Create New](#)

New Grouping

[Save](#) [Delete](#) [Cancel](#)

**Name:**

**Viewed:** ☒

Parameters [Create New](#)

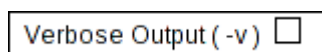
- Now, click **Create New** next to Parameters, and then click the **Edit** link next to "New Parameter". Below is a description of each field required to define a parameter. After describing these fields in general, we will demonstrate how they were used to add parameters for the mutate tool.
  - **Logical Name:** As with Inputs and Outputs before, the logical name is the name used for this parameter within BioExtract Server.
  - **Physical Name:** This is the actual parameter name, as it should appear on the command line. Note that BioExtract Server will not add a "-" automatically, so please remember to do so if your tool requires a "-" before the parameter name.
  - **Description:** Optional; Additional information about the parameter.

- **Parameter Type:** This specifies the way values are given for the parameter. Depending on the type, they may be entered directly by the user or chosen from a list of pre-defined values. Below are the details for each type:

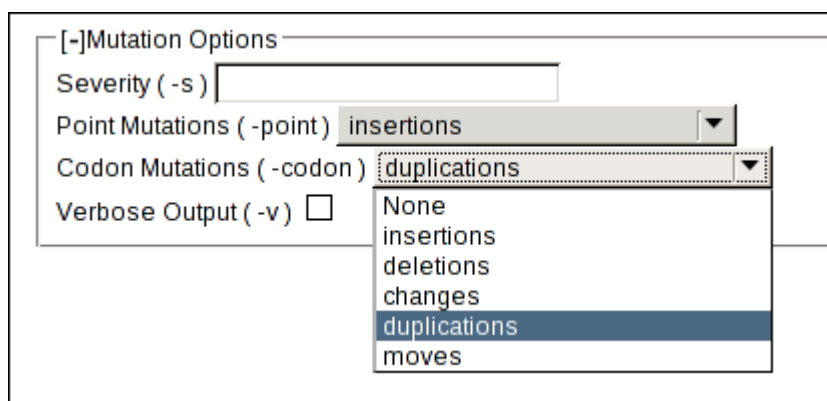
- **text:** Creates a field where the value can be typed in directly. Also useful for numeric values.

A screenshot of a text input field. The label "Severity (-s)" is on the left, and an empty rectangular box is on the right for typing.

- **checkbox:** Creates a checkbox. Useful for parameters that don't have any additional values (for example, some programs will print extra information if a -v is present on the command line). If checked, the parameter (specifically, the **Physical Name**) will appear on the command line.

A screenshot of a checkbox input field. The label "Verbose Output (-v)" is on the left, followed by an unchecked square checkbox.

- **select:** Creates a "drop-down" menu from which one of several possible values can be chosen.

A screenshot of a software interface titled "[ - ] Mutation Options". It contains three rows: "Severity (-s)" with a text box, "Point Mutations (-point)" with a drop-down menu showing "insertions", and "Codon Mutations (-codon)" with a drop-down menu showing "duplications". Below these is a "Verbose Output (-v)" checkbox. The "Codon Mutations" drop-down menu is open, showing a list of options: "None", "insertions", "deletions", "changes", "duplications" (which is highlighted), and "moves".

When using the "select" type, the possible values must be defined. This is not required for any of the other types. To define a set of values, click on the **Create New** link next to "Values". Then, click **Edit** next to the "New Parameter Value". "Value" is the value as it should appear on the command line. If "Is Default" is checked, this value will be selected by default when the tool's interface is shown.

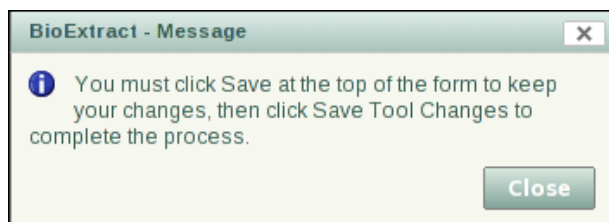
- **textarea:** Identical to "text" type above.
- **Tab Order:** Optional; If you would like the parameters to be displayed in a certain order, this number gives the rank of this particular parameter. The ordering is from lowest-to-highest, so a parameter with a tab order of 2 appears above one with an order of 3.
- **Is Mandatory:** If checked, the tool will not be able to execute unless a value for this parameter has been entered.
- Click the **Save** link under "New Parameter".

This is the completed "-db" parameter for Fetch Records:

The screenshot shows a web interface for defining a new parameter. The main heading is 'Parameters' with a 'Create New' link. Below it is a 'New Parameter' section. At the top of this section are 'Save', 'Delete', and 'Cancel' links. The form contains the following fields: 'Logical Name' with the value 'Database', 'Physical Name' with the value '-db', and a 'Description' text area containing 'Name of database containing desired sequences'. Below these is a 'Parameter Type' dropdown menu set to 'select', a 'Tab Order' field with the value '0', and an 'Is Mandatory' checkbox which is checked. At the bottom of the parameter form is a 'Values' section with a 'Create New' link. It lists two values: 'ncbi' and 'embl', each with an 'Edit' link. At the very bottom of the interface is a green 'Save Tool' button.

### Step 7: Saving the Local Tool


Once all of the necessary inputs, outputs, and parameters have been defined, click on the **Save Tool** button at the bottom of the screen. Please note that all of the "sub forms" opened for each input, output, and parameter must be saved before clicking **Save Tool**. If any of the sub forms are still open, the following error message will be displayed:



## Running a Local Tool

Once the tool is saved, it will appear under the "My Tools" group in the "Available Tools" menu.

When you select the tool, the interface presented is the same one used by all BioExtract Server tools.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

QueryExtractsToolsWorkflowsGroupsHelp

Demo Workflow

Available Tools

Tools

Alignment Tools

BioMart

CpGAT

Edit Tools

Information Tools

My Tools

B Find

EBI\_InterproScan

Fetch Records

Format conversion

backtranseq

bget

cpgplot

embosdata

emma

fconsense

fneighbor

fseqboot

Add a New Tool

Customize Your Tools

### Fetch Records

This tool takes a file containing a list of ids and the name of a data source (i.e. ncbi, embl, refseq). The tool accesses the data source and retrieves the records associated with the ids.. No help available.

#### Input Data

Ids

☐ Use records on Extracts page formatted as FASTA

☐ Use previously executed tool results Select tool Select result

☐ Upload data saved on your computer

Choose File No file chosen

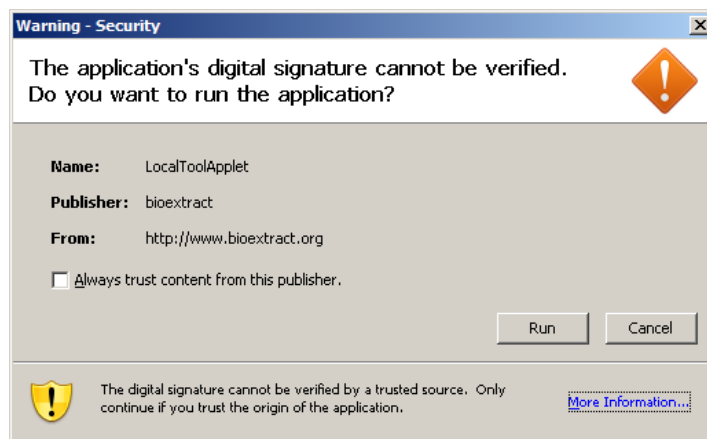
☒ Paste or type data into the text area

#### Parameter Settings

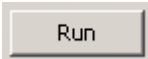
[ - ]Required

Database ( -db ) ncbi

A few moments after clicking "Execute", a popup window should appear, followed by another one that looks like this:



This is normal. A Java applet is used to execute the local tool, and applets cannot execute programs or write files without permission, which can only be granted if the applet is placed in a digitally signed .jar file. Since we have signed the file ourselves without using a certificate from one of the third-party Certificate Authorities, the browser reports that the signature cannot be verified.

Once you click the  button, the applet will download the input files from the server, execute the tool, and upload the output files, displaying a short message for each step. Once it is finished, you may close the applet window.

## BIOEXTRACT SERVER WORKFLOWS

### Introduction

Users do not explicitly create workflows in the BioExtract Server, but implicitly do so by working with the system. As you work in the BioExtract Server, all of your tasks (such as executing queries against selected data sources, applying analytic tools, and saving data extracts) are recorded. At any point, a workflow comprising the performed set of tasks can be saved and subsequently executed as a single unit. Individual tasks within the workflow may also be modified or deleted by the workflow owner.

To illustrate a BioExtract Server workflow, consider the task of carrying out a phylogenetic analysis for a set of proteins where the starting point is a particular genomic coding sequence representing only one member of the gene family in a given species. In the BioExtract Server, the steps for accomplishing this task involve:

- Selecting the *tblastx* tool and providing the accession number of the nucleotide sequence record as input.
- The output from *tblastx*, a BLAST report along with a set of records representing similar sequences, is parsed using a formatting template to produce an initial data extract (i.e. a set of matching nucleotide sequences).
- The resulting data extract is saved.
- The resulting data extract is used as input into the tool *Vmatch* (see <http://www.vmatch.de/>) to remove duplicate sequences.
- The tool *fetchTranslation* is run. This tool is defined to use the current data extract as input (in GenBank format) and returns the protein translations from the GenBank-annotated coding sequence (CDS) regions (in FASTA format).
- The *ClustalW* tool is run to create the multiple sequence alignment with the input specified as coming from the previously executed tool (*fetchTranslation*) and to define and draw a dendrogram that represents how the sequences are related.
- The *TCoffee* tool is run to create the multiple sequence alignment with the input specified as coming from the previously executed tool (*fetchTranslation*) and to define and draw a dendrogram that represents how the sequences are related.



## Creating a Workflow


As an example, let's create the workflow outlined above.

### Step 1: Preparing to create the workflow

If you have just signed into the BioExtract Server, it's not necessary to "clear" the workflow memory. But if you have been working with the system and want to begin creating a workflow, it is necessary to clear any previously executed tasks from memory.

**To clear previously executed tasks from memory:**

1. Click the Workflows tab and select the **Create and Import Workflows** heading in the Workflows tree. The "Create and Import Workflows" form opens in the right panel.
2. In the Create Workflow section, click the **Record Workflow** button. This will erase any previously executed tasks.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

QueryExtractsTools**Workflows**GroupsHelp

Demo Workflow

**Workflows**

- Create and Import Workflows
- + Biomart SNP
- + Blastn-ClustalW Example
- + Carol's format conversion
- + CpGAT Maize
- + Gene function prediction in Macaca Fasciculus
- + Liliopsida petb protein alignment
- + Multiple Protein Alignment Profiling
- + NP\_000135 Multiple Sequence Alignment
- + NP\_000135 Multiple Sequence Alignment
- + Nucleotide InterProScan
- + Phylogenetic analysis from a gene query
- + Protein alignment from a gene query
- + Query- align
- + SH2 domain (in homo) screen
- + Sequence Format Conversion Example
- + Test Copy
- + Unique Asparagus petb nucleotides
- + mRNA Markup
- + tblastx non-redundant alignment

**Create and Import Workflows**

**Create Workflow**

Record Workflow

**Save Workflow**

Name:


Description:

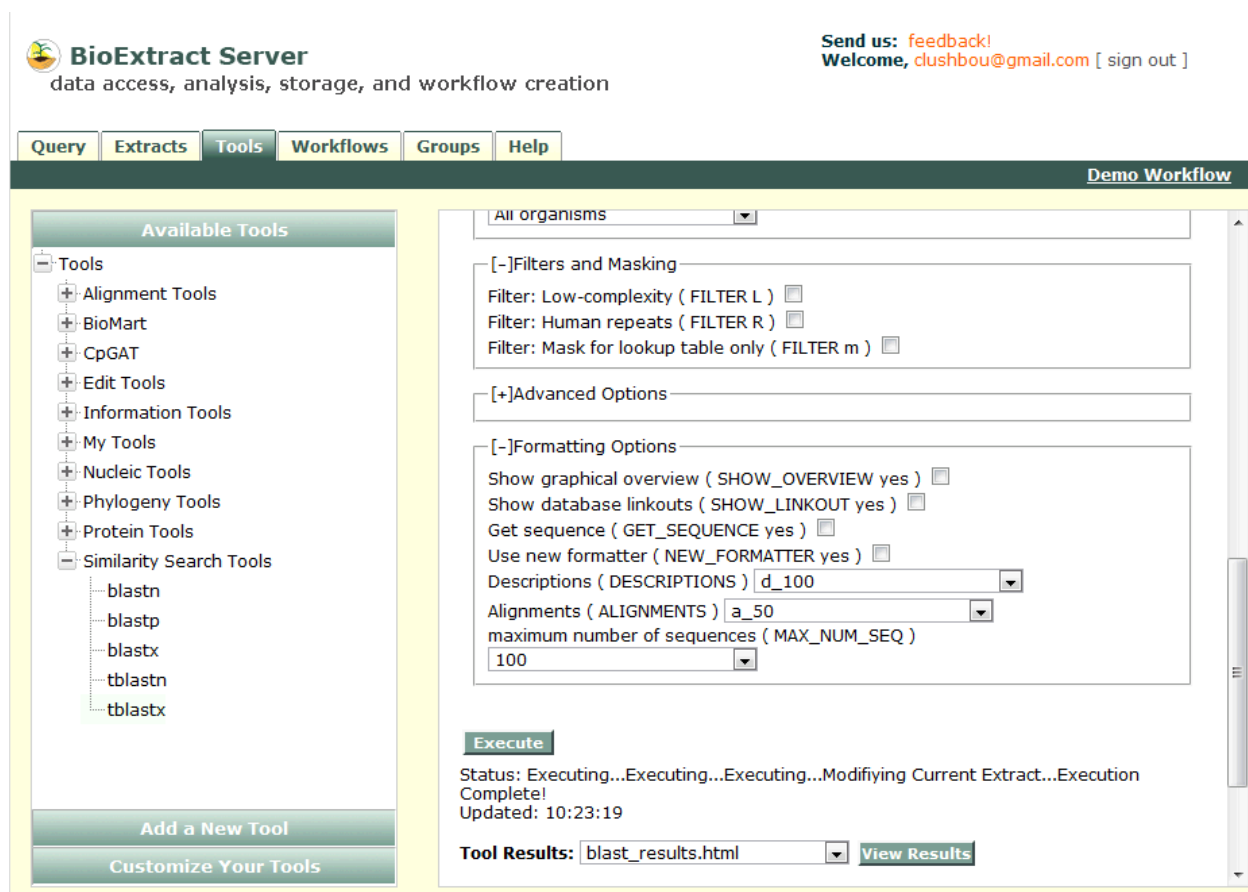
Save

**Import Workflow**

No file chosen

## Step 2: Execute the *tblastx* tool from the Tools tab

1. Click the Tools tab. In the Available Tools menu expand the  **Similarity Search Tools** group and select the *tblastx* tool. The tool's form opens in the right panel.
2. In the Input Data section, specify **Paste or type data into the text area** for the input source and enter the accession number of the nucleotide sequence record as input, "L16896" (without the quotes).



The screenshot displays the BioExtract Server web interface. At the top, the header includes the BioExtract Server logo, the tagline "data access, analysis, storage, and workflow creation", and user information: "Send us: feedback! Welcome, clushbou@gmail.com [ sign out ]". Below the header is a navigation bar with tabs: Query, Extracts, Tools, Workflows, Groups, and Help. The "Tools" tab is active, and a "Demo Workflow" link is visible on the right.

The main content area is divided into two panels. The left panel, titled "Available Tools", shows a tree structure of tool categories. The "Similarity Search Tools" category is expanded, listing tools: blastn, blastp, blastx, tblastn, and tblastx. The "tblastx" tool is selected and highlighted in green. Below this list are buttons for "Add a New Tool" and "Customize Your Tools".

The right panel displays the configuration form for the selected *tblastx* tool. It includes a dropdown for "All organisms". Below this are sections for "Filters and Masking" (with checkboxes for "Filter: Low-complexity ( FILTER L )", "Filter: Human repeats ( FILTER R )", and "Filter: Mask for lookup table only ( FILTER m )"), "Advanced Options" (collapsed), and "Formatting Options" (expanded). The "Formatting Options" section contains several settings: "Show graphical overview ( SHOW\_OVERVIEW yes )" (checkbox), "Show database linkouts ( SHOW\_LINKOUT yes )" (checkbox), "Get sequence ( GET\_SEQUENCE yes )" (checkbox), "Use new formatter ( NEW\_FORMATTER yes )" (checkbox), "Descriptions ( DESCRIPTIONS )" (dropdown set to "d\_100"), "Alignments ( ALIGNMENTS )" (dropdown set to "a\_50"), and "maximum number of sequences ( MAX\_NUM\_SEQ )" (input field set to "100").

At the bottom of the right panel, there is an "Execute" button. Below it, the status is shown as "Status: Executing...Executing...Executing...Modifying Current Extract...Execution Complete!" with a timestamp "Updated: 10:23:19". The "Tool Results" section shows a dropdown set to "blast\_results.html" and a "View Results" button.

3. Click the Execute button. The output from the execution of the *tblastx* tool is a blast report; furthermore the *tblastx* report is turned into a data extract viewable on the Extracts page.

View File /usr/local/BioStreamServer/tmpFiles/tblastx/1309274312876

[\[Download This File\]](#)

[\[Download/View This PDF File\]](#)

TBLASTX 2.2.25+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: OKW69XE7012

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)  
14,157,082 sequences; 36,508,476,836 total letters

Query= MUSCOL18A Mus musculus collagen alpha 1 type XVIII mRNA, 5'end. 2805 bp  
Length=2805

Sequences producing significant alignments:


	(Bits)	Value	N
gi 1167902 gb U34612.1 MMCOL18A07 Mus musculus alpha-1(XVIII)...	155	0.0	1
gi 194226339 ref XR_035837.2  PREDICTED: Equus caballus misc...	181	2e-179	1
gi 37606039 emb AL954258.2  Pan troglodytes chromosome 22 clo...	346	3e-175	1
gi 226958461 ref NG_011903.1  Homo sapiens collagen, type XVI...	343	4e-174	1
gi 66773438 gb AC159334.9  Mus musculus 10 BAC RP24-369D19 (R...	433	3e-168	1
gi 29500747 emb BX322561.1  Homo sapiens chromosome 21 from P...	343	4e-167	1
gi 7717445 emb AL163302.2  Homo sapiens chromosome 21 segment...	343	3e-159	1
gi 45383787 ref NM_204164.1  Gallus gallus collagen, type XVI...	284	2e-146	1

Query Extracts Tools Workflows Groups Help

[Demo Workflow](#)








Query returned **50** records.

Databases

Database	Total	Database Description
 NCBI	50	


Select Records Export Records Save Extract


Results Page: << first < prev 1 next > last >>

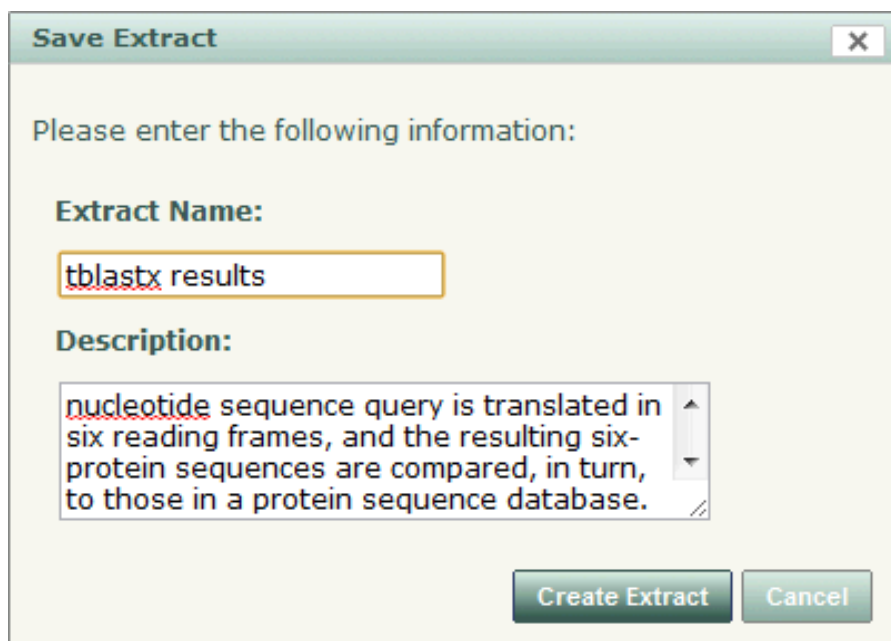
External Link	Local Details	Description
 <a href="#">NM_053489</a>	<a href="#">view record</a>	rattus norvegicus collagen, type xviii, alpha 1 (col18a1), mrna
 <a href="#">BC064817</a>	<a href="#">view record</a>	mus musculus collagen, type xviii, alpha 1, mrna (cdna clone
 <a href="#">BC043697</a>	<a href="#">view record</a>	mus musculus procollagen, type xviii, alpha 1, mrna (cdna clone
 <a href="#">BC066080</a>	<a href="#">view record</a>	mus musculus collagen, type xviii, alpha 1, mrna (cdna clone
 <a href="#">D17546</a>	<a href="#">view record</a>	muscollag mus musculus mrna for collagen, partial cds
 <a href="#">L16898</a>	<a href="#">view record</a>	muscol18a mus musculus collagen alpha 1 type xviii mrna, 5'end
 <a href="#">NM_001109991</a>	<a href="#">view record</a>	mus musculus collagen, type xviii, alpha 1 (col18a1), transcript

### Step 3: Save the resulting data extract

1. Click the Extracts tab. The tblastx data extract displays.


2. Click the  button. The “Save Extract” dialog box opens.

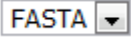
3. Enter a name and description for the data extract and click the  button. The data extract is saved and becomes a searchable data source available on the Query page In Available Data Sources under the heading Miscellaneous.




The image shows a 'Save Extract' dialog box with a title bar containing a close button (X). The main text says 'Please enter the following information:'. There are two sections: 'Extract Name:' with a text input field containing 'tblastx results', and 'Description:' with a text area containing 'nucleotide sequence query is translated in six reading frames, and the resulting six-protein sequences are compared, in turn, to those in a protein sequence database.' At the bottom right are two buttons: 'Create Extract' and 'Cancel'.

### Step 4: Remove duplicate sequences in the data extract using the *xmknr* tool



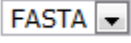
1. Click the Tools tab. In the Available Tools menu expand the  Edit Tools group and select the *xmknr* tool. The tool’s form opens in the right panel.


2. In the Input Data section, specify ☒ Use records on Extracts page formatted as  for the input source.

3. In the Parameter Settings section, set Sequence type (protein or DNA) ( -a )  to dna.

4. Click the Execute button. A number of result files are created. In addition, the data extract visible on the Extracts page has been modified to remove any duplicate sequence records.

**Step 5: Convert the data extract to GenBank format using the *FormatConversion* tool (as the next tool requires GenBank format to run)**

1. Click the Tools tab. In the Available Tools menu expand the  **Edit Tools** group and select the *FormatConversion* tool. The tool's form opens in the right panel.
2. In the Input Data section, specify  **Use records on Extracts page formatted as**  for the input source.
3. In the Parameter Settings section, set **To Format** to *genbank* and **From Format** to *fasta*.
4. Click the Execute button. The records in the data extract are now converted to *genbank* format.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ [sign out](#) ]

**Query** **Extracts** **Tools** **Workflows** **Groups** **Help**

**Demo Workflow**

**Available Tools**


- Tools
  - Alignment Tools
  - BioMart
  - CpGAT
  - Edit Tools
    - FormatConversion**
    - xmknr
  - Information Tools
  - My Tools
  - Nucleic Tools
  - Phylogeny Tools
  - Protein Tools
  - Similarity Search Tools



[Add a New Tool](#)  
[Customize Your Tools](#)


Convert from one allowed format to another allowed format. [More Information.](#)

**Input Data**

Input File

☒ Use records on Extracts page formatted as 


☐ Use previously executed tool results  


☐ Upload data saved on your computer  
 No file chosen


☐ Paste or type data into the text area

**Parameter Settings**


[ - ] Conversion Options

To Format ( -of ) 

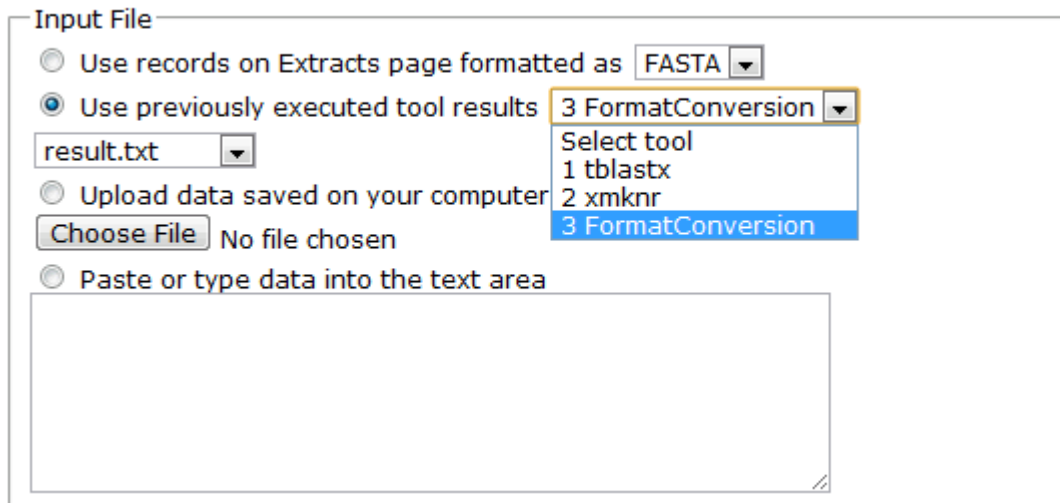
From Format ( -if ) 



**Step 6: Retrieve the protein translations from the CDS regions of the DNA sequences using the *FetchTranslation* tool**

1. From the Tools page, in the Available Tools menu expand the  **Information Tools** group and select the *FetchTranslation* tool. The tool's form opens in the right panel.
2. In the Input Data section, specify **Use previously executed tool results** for the input source and select *FormatConversion* and *result.txt* from the associated drop-down menus.
3. Click the Execute button. The tool *FetchTranslation* runs and returns the protein translations in FASTA format.

### Input Data



Input File

☐ Use records on Extracts page formatted as FASTA ▾

☒ Use previously executed tool results 3 FormatConversion ▾

result.txt ▾


☐ Upload data saved on your computer Choose File No file chosen

☐ Paste or type data into the text area

Select tool

- 1 tblastx
- 2 xmknr
- 3 FormatConversion

## Step 7: Create a *ClustalW2* multiple sequence alignment and dendrogram with the protein translations retrieved by *FetchTranslation*

1. From the Tools page, in the Available Tools menu expand the  **Alignment Tools** group and select the *ClustalW2* tool. The tool's form opens in the right panel.
2. In the Input Data section, specify **Use previously executed tool results** for the input source and select *FetchTranslation* and *fetchTranslation\_results.txt* from the associated drop-down menus.
3. Click the Execute button. The tool *ClustalW2* runs and creates a multiple sequence alignment and draws a dendrogram that represents how the sequences are related.

## ClustalW2

Multiple Sequence Alignments. [More Information.](#)

### Input Data

Input file

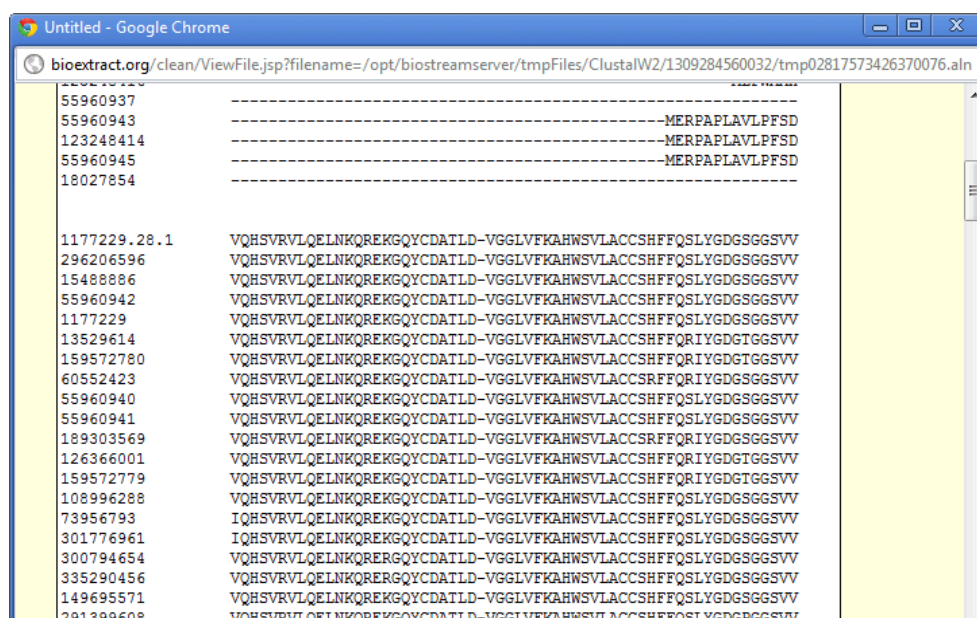
☐ Use records on Extracts page formatted as FASTA

☒ Use previously executed tool results 4 FetchTranslation


fetchTranslation\_results.txt

☐ Upload data saved on your computer  
 No file chosen

☐ Paste or type data into the text area



## Step 8: Create a *TCoffee* multiple sequence alignment and dendrogram with the protein translations retrieved by *FetchTranslation*

1. From the Tools page, select the *TCoffee* tool in the  **Alignment Tools** group. The tool's form opens in the right panel.
2. In the Input Data section, specify **Use previously executed tool results** for the input source and select *FetchTranslation* and *fetchTranslation\_results.txt* from the associated drop-down menus.
3. Click the Execute button. The tool *TCoffee* runs and creates a multiple sequence alignment and draws a dendrogram that represents how the sequences are related.

### TCoffee

Computes a multiple sequence alignment and the associated phylogenetic tree for Protein, RNA and DNA sequences. [More Information.](#)

#### Input Data

Input sequences (FASTA format)

☐ Use records on Extracts page formatted as FASTA

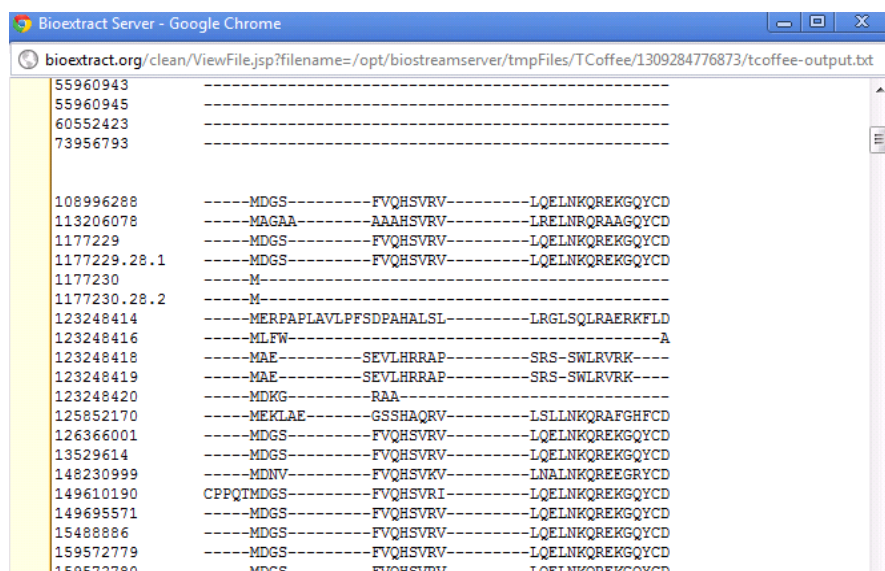
☒ Use previously executed tool results 4 FetchTranslation

fetchTranslation\_results.txt

☐ Upload data saved on your computer

Choose File No file chosen

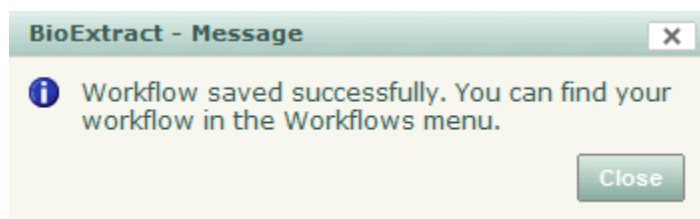
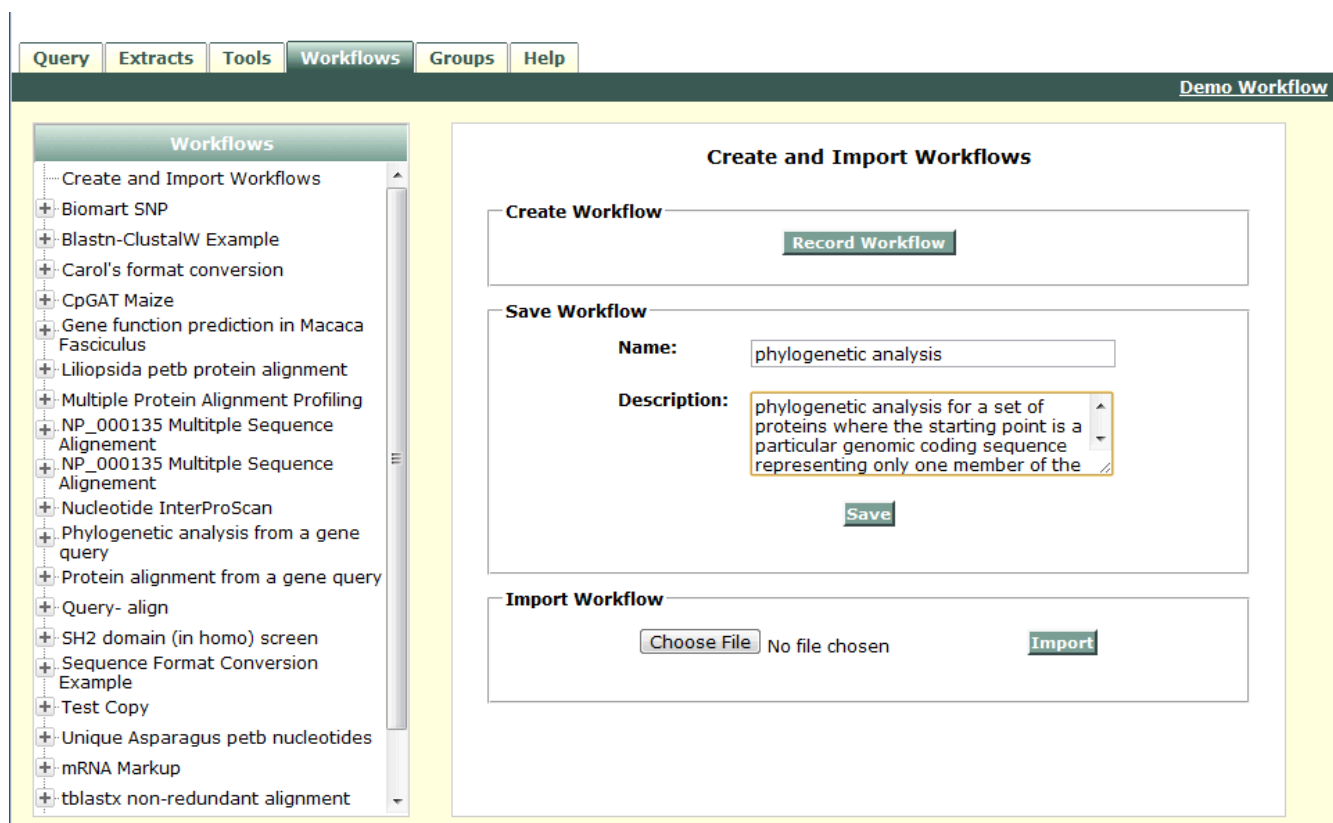
☐ Paste or type data into the text area






**Step 9: Save the workflow**

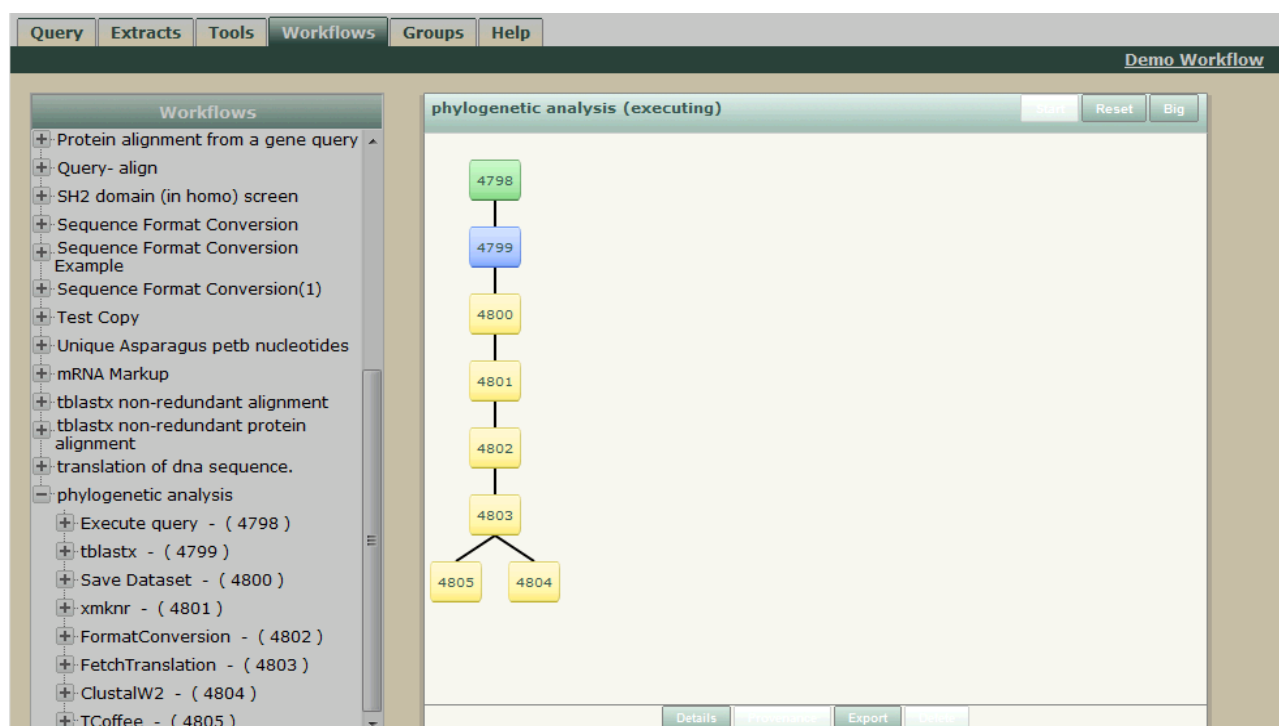
1. Click the Workflows tab and select the **Create and Import Workflows** heading in the Workflows tree. The “Create and Import Workflows” form opens in the right panel.
2. In the Save Workflow section, enter a name and description for the new workflow, and then click the **Save** button. The workflow is saved and displays in the Workflows menu.



## Executing a Workflow

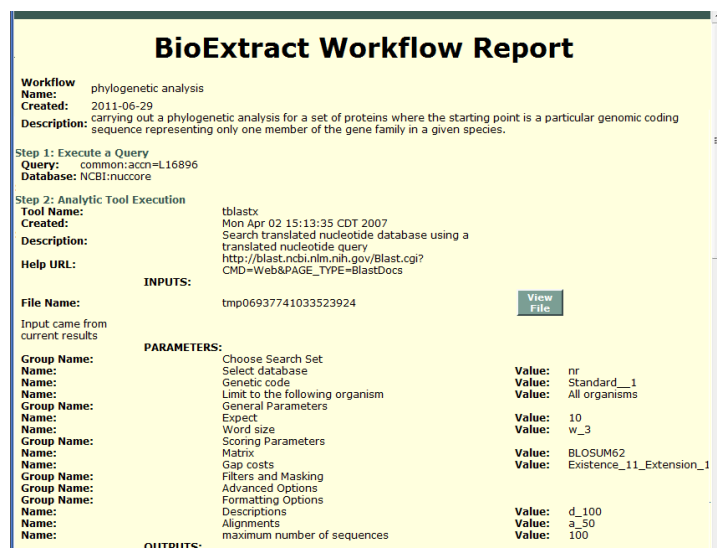
To execute a workflow, select the desired workflow from the Workflows menu on the Workflows page. A graphical representation of the workflow displays in the right panel.

Click the  button at the top of the workflow panel. When a step within the workflow begins to run, it will turn blue. When the execution completes, the color of the step will change to green.



The screenshot shows the BioExtract Workflows interface. The top navigation bar includes tabs for Query, Extracts, Tools, Workflows, Groups, and Help. The 'Workflows' tab is active, displaying a list of workflows on the left and a detailed view of the 'phylogenetic analysis (executing)' workflow on the right. The workflow diagram shows a sequence of steps: 4798 (green), 4799 (blue), 4800 (yellow), 4801 (yellow), 4802 (yellow), 4803 (yellow), 4804 (yellow), and 4805 (yellow). The 'Start' button is visible at the top right of the workflow panel.

When execution completes, the **Provenance** button becomes enabled. Click the **Provenance** button to see the provenance information associated with the execution of the workflow.



The screenshot displays the 'BioExtract Workflow Report' for the 'phylogenetic analysis' workflow. The report includes the following information:

- Workflow Name:** phylogenetic analysis
- Created:** 2011-06-29
- Description:** carrying out a phylogenetic analysis for a set of proteins where the starting point is a particular genomic coding sequence representing only one member of the gene family in a given species.
- Step 1: Execute a Query**
  - Query:** common:acon-L16896
  - Database:** NCBI:nucleotide
- Step 2: Analytic Tool Execution**
  - Tool Name:** tblastx
  - Created:** Mon Apr 02 15:13:35 CDT 2007
  - Description:** Search translated nucleotide database using a translated nucleotide query
  - Help URL:** [http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs)
- File Name:** tmp06937741033523924 [View File](#)
- Input:** Input came from current results
- PARAMETERS:**
  - Group Name:** Choose Search Set
  - Name:** Select database
  - Name:** Genetic code
  - Name:** Limit to the following organism
  - Group Name:** General Parameters
  - Name:** Expect
  - Name:** Word size
  - Group Name:** Scoring Parameters
  - Name:** Matrix
  - Name:** Gap costs
  - Group Name:** Filters and Masking
  - Group Name:** Advanced Options
  - Group Name:** Formatting Options
  - Name:** Descriptions
  - Name:** Alignments
  - Name:** maximum number of sequences
- OUTPUTS:**
  - Value:** nr
  - Value:** Standard\_1
  - Value:** All organisms
  - Value:** 10
  - Value:** w\_3
  - Value:** BLOSUM62
  - Value:** Existence\_11\_Extension\_1
  - Value:** d\_100
  - Value:** a\_50
  - Value:** 100


## Modifying a Workflow


One of the major advantages of BioExtract Server workflows is they can be modified and executed using different data and parameter settings.

**Warning:** Workflow modifications are saved permanently if you are the owner of the workflow. Otherwise, they are saved temporarily. Before you modify a workflow, you may want to make a copy, so you don't overwrite the original workflow. See “Copying a Workflow” for more information.

As an example, assume the original workflow contained a query step using accession number “L16896” and you’re interested in running the same workflow using accession number “NM\_005341”. You would also like to change some of the *tblastx* tool parameter settings.

### To modify a query and tool step in a workflow:

1. From the Workflows page, select the workflow you want to modify (i.e.  **phylogenetic analysis** ). This will show you all of the steps included in that workflow. Don't forget to sign in to see your workflows.

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

QueryExtractsTools**Workflows**GroupsHelp

Demo Workflow

Workflows

+ Protein alignment from a gene query

+ Query- align

+ SH2 domain (in homo) screen

+ Sequence Format Conversion

+ Sequence Format Conversion Example

+ Sequence Format Conversion(1)

+ Test Copy

+ Unique Asparagus petb nucleotides

+ mRNA Markup

- phylogenetic analysis

+ Execute query - ( 4798 )

+ tblastx - ( 4799 )

+ Save Dataset - ( 4800 )

+ xmknr - ( 4801 )

+ FormatConversion - ( 4802 )

+ FetchTranslation - ( 4803 )

+ ClustalW2 - ( 4804 )

+ TCOffee - ( 4805 )

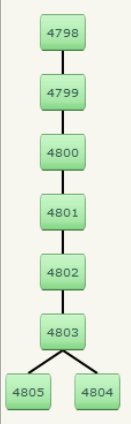
+ tblastx non-redundant alignment

+ tblastx non-redundant protein alignment

+ translation of dna sequence.

phylogenetic analysis (complete)

StartResetBig



```
graph TD; 4798 --> 4799; 4799 --> 4800; 4800 --> 4801; 4801 --> 4802; 4802 --> 4803; 4803 --> 4804; 4803 --> 4805;
```

DetailsProvenanceExportDelete

Page 43

If you want to change the query request, click the plus sign to expand the “Execute query” step. The step will expand and display modifiable sub-steps (Query and Databases). Select the Query heading. A Query Information form opens in the right panel. You can change the search field and the search term. Change the search term “L16896” to NM\_005341 and click the Save button to keep your changes. Table 1 displays valid search terms that may be used when modifying a query.

### Query Information

**Properties**

**Query:**

[Query examples.](#)

### Query Information

**Properties**

**Query:**

[Query examples.](#)

Field	Definition	Search Term	Search Term Examples
<b>Accession</b>	Contains the unique accession number of the sequence or record	common:accn	common:accn=NM_005341
<b>All Text</b>	Contains all terms from all searchable database fields in the database.	common:all	common:all='BTB domain'
<b>Author</b>	Contains all authors from all references in the database records	common:author	common:author=Zhang
<b>Definition</b>	Includes only those words found in the definition line of a record.	common:defn	common:defn='Homo sapiens' AND common:defn='BTB domain' AND common:defn=' zinc finger '
<b>Feature Key</b>	Contains the biological features assigned or annotated to the nucleotide sequences and defined in the DDBJ/EMBL/GenBank	common:fkey	common:fkey=gene

	Feature Table		
<b>Gene</b>	Contains the standard and common names of genes found in the database records	common:gene	common:gene=Zbtb16
<b>Id</b>		common:id	common:id=157694498
<b>Keywords</b>	Contains special index terms from the controlled vocabularies associated with the GenBank, EMBL, DDBJ, SWISS-Prot, PIR, PRF, or PDB databases.	common:keyword	common:keyword= Zbtb8
<b>Species</b>		common:species	common:species= 'Camphor tree'
<b>Taxonomy</b>	Contains the scientific and common names for the organisms associated with protein and nucleotide sequences.	common:taxonomy	common:taxonomy='Mus musculus'
<b>Title</b>	Title of the journal abbreviation	Common: title	Common: title=" Plant Physiol"

**Table 1.** List of valid search fields and example search terms.

If you want to change *tblastx* tool parameters, click the plus sign to expand the “tblastx” tool step. The step will expand and display the modifiable sub-step Tool. Select the Tool heading. The tool form opens in the right panel. Change parameters according to your preferences. Input into a tool may be modified if the original data was entered in the text box or uploaded as a file. Click the Save button to keep your changes.

The screenshot displays the NCBI Bioinformatics Workflows web application. The top navigation bar includes tabs for Query, Extracts, Tools, Workflows (selected), Groups, and Help. A 'Demo Workflow' button is located on the right. The left sidebar shows a tree view of workflows, with 'tblastx non-redundant alignment' expanded to show sub-steps: 'Execute query - ( 4830 )', 'tblastx - ( 4831 )', and 'Tool'. The 'Tool' sub-step is selected, opening a configuration panel on the right. This panel contains several sections: 'Limit to the following organism ( ENTREZ\_QUERY )' with a dropdown set to 'All organisms'; '[-]Filters and Masking' with checkboxes for 'Filter: Low-complexity ( FILTER L )', 'Filter: Human repeats ( FILTER R )', and 'Filter: Mask for lookup table only ( FILTER m )'; '[-]Advanced Options' with a text input for 'Other advanced ( OTHER\_ADVANCED )'; and '[-]Formatting Options' with checkboxes for 'Show graphical overview ( SHOW\_OVERVIEW yes )', 'Show database linkouts ( SHOW\_LINKOUT yes )', 'Get sequence ( GET\_SEQUENCE yes )', and 'Use new formatter ( NEW\_FORMATTER yes )'. It also includes dropdowns for 'Descriptions ( DESCRIPTIONS )' (set to 'd\_50'), 'Alignments ( ALIGNMENTS )' (set to 'a\_50'), and a text input for 'maximum number of sequences ( MAX\_NUM\_SEQ )' (set to '50'). At the bottom of the panel is a 'Save' button. Below the configuration panel, the 'Tool Results' section shows 'blast\_results.html' with a 'View Results' button.

## Copying a Workflow

**To copy a workflow:**

1. Click the **Create and Import Workflows** heading on the Workflows page. The “Create and Import Workflows” form opens in the right panel.
2. In the Create Workflow section, click the **Record Workflow** button.
3. Next, open the workflow you want to make a copy of and run it. The BioExtract Server will record the running workflow in the background.
4. Once execution is complete, click the **Create and Import Workflows** heading on the Workflows page. The “Create and Import Workflows” form opens in the right panel.
5. In the Save Workflow section, enter a new Name and keep or change the Description, then click the Save button. The name of the new “copied” workflow will appear in the Workflows menu. You can now modify the original workflow or the copy.

**Save Workflow**

**Name:**

**Description:**

**Save**

## BIOEXTRACT SERVER GROUPS

### About Groups

Groups provide a collaborative environment to facilitate the sharing of data extracts, analytic tools and workflows.

Registered users can create new groups by using the Create Group option on the Groups page. When you create a group, you'll be able to invite others to join. Plus you'll be able to share your tools, workflows and data extracts with others. In addition, you'll be able to modify your shared workflows, and remove tools, data extracts and workflows from group access. Groups you own are listed under Owned Groups.

**BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

Query Extracts Tools Workflows **Groups** Help

[Demo Workflow](#)

**Owned Groups**  
A New Group  
BRIN  
USD Bioteam

**Member Groups**  
Place Holder  
Users

**Additional Actions**  
Create Group

By creating groups and inviting other BioExtract Server users to participate in them, you can easily share analytic tools, workflows, and data extracts. To create a new group, ensure you are logged in and click Create Group on the left. [Read more about creating groups here.](#)

Groups you own are listed under Owned Groups, and groups of which you are a member are listed under Member Groups. As a member, you can view and execute any of the tools, workflows, and data extracts that you've been given permission to use.

Groups of which you are a member are listed under Member Groups. As a group member, you can view and execute any of the tools, workflows and data extracts that you've been given permission to use.

To use group tools, sign in and click the Tools tab. In Available Tools, expand **My Tools**. By default, group tools are listed along with your private tools. Locate and select the group tool you want to use. The tool form will open in the right panel.

To use group workflows, sign in and click the Workflows tab. By default, all workflows (public, private and group) are given in one list. Locate and select the group workflow you want to use. The workflow graph and its control buttons will open in the right panel.

Group data extracts can be found on the Query page in Available Data Sources under the Miscellaneous heading.

## Creating a Group

The BioExtract Server provides a facility to create groups of registered users. Groups allow registered users to share tools, workflows and data extracts.

### To create a group:

1. You'll first need to create a user account and sign in. At the top right corner of any BioExtract Server page, select **register**. Fill in the required fields and click **Add User**. Your account will be created immediately. Return to the BioExtract Server and sign in by clicking the **sign in** link at the top right corner of any page.
2. Click the **Groups** tab.
3. In the left panel, under the **Additional Actions** heading, click **Create Group**. A new group form opens in the right panel.
4. Click **Edit** to the right of the heading "A New Group." Enter a Name and Description for this new group, then click the **Save** button. A new group is created. You'll see your new group under the **Owned Groups** heading in the left panel.

The "Name and Description" of the newly created group can be edited by selecting the group and editing the information. Once the information has been edited, click the **Save** link at the top of the panel to keep your changes.

The screenshot displays the BioExtract Server interface. At the top, there is a navigation bar with tabs: Query, Extracts, Tools, Workflows, Groups, and Help. The 'Groups' tab is currently selected. On the right side of the navigation bar, there is a 'Demo Workflow' link. The main content area is divided into two panels. The left panel, titled 'Owned Groups', contains a list of groups: 'A New Group' (with a checkmark), 'BRIN', and 'USD Bioteam'. Below this, there is a section for 'Member Groups' with 'Place Holder' and 'Users'. At the bottom of the left panel, under 'Additional Actions', is a 'Create Group' button. The right panel shows the 'A New Group' form, which has tabs for 'A New Group', 'Members', 'Tools', 'Workflows', and 'Extracts'. The 'A New Group' tab is active. Inside this tab, there are buttons for 'Save', 'Delete', and 'Cancel'. Below these buttons, there is a 'Name' field with the text 'Sanford Research' and a 'Description' field with the text 'Sanford Children's Hospital Research Group'.



## Adding Tools, Workflows and Data Extracts to a Group

After you create a group, you can add elements, such as analytic tools, data extracts and workflows to your group.

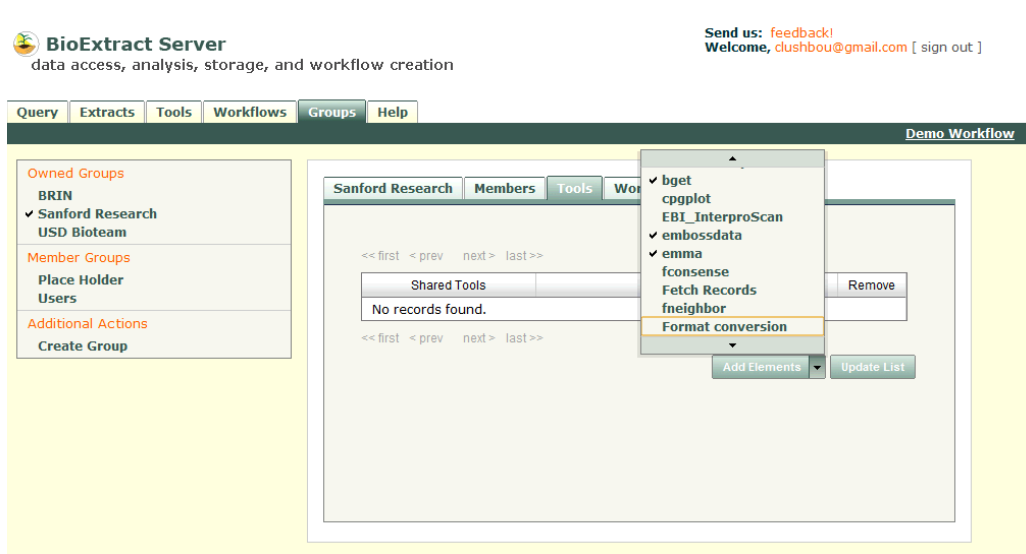
**Note:** Before adding tools, data extracts and workflows to your group, you must first add them to your private BioExtract Server account.


**To add tools, data extracts and workflows to a group:**

1. Sign in to the BioExtract Server and click the Groups tab.
2. In the left panel, under the **Owned Groups** heading, select the name of the group that you want to add an element to. The group form opens in the right panel.
3. Click the Tools, Workflows or Extracts tab depending on what you'd like to add. The Add Elements form displays.



4. Click the black arrow on the **Add Elements** button. A list of tools, workflows and data extracts appears. Select the name of the element that you'd like to add. Choose as many elements as you like, then click the **Add Elements** button. The selected element will be added to the form.



**BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

QueryExtractsToolsWorkflowsGroupsHelp

Demo Workflow

Owned Groups

BRIN

✓ Sanford Research

USD Bioteam

Member Groups

Place Holder

Users

Additional Actions

Create Group

Sanford ResearchMembersToolsWorkflowsExtracts

<< first < prev 1 next > last >>

Shared Tools	Description	Remove
bget	<a href="http://soap.genome.jp/">http://soap.genome.jp/</a>	<input type="checkbox"/>
embosdata	Find and retrieve EMBOSS data files	<input type="checkbox"/>
emma	Multiple sequence alignment (ClustalW wrapper)	<input type="checkbox"/>

<< first < prev 1 next > last >>

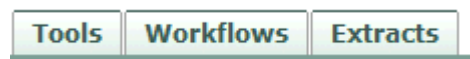
Add ElementsUpdate List

## Removing Tools, Workflows and Data Extracts from a Group

If you're the owner of a group, you can remove tools, data extracts and workflows from that group.


**To remove tools, data extracts and workflows from a group:**

1. Sign in to the BioExtract Server and click the Groups tab.
2. In the left panel, under the **Owned Groups** heading, select the name of the group you want to remove an element from. The group form opens in the right panel.
3. Click the Tools, Workflows or Extracts tab depending on what you'd like to remove. The Add Elements form displays.



4. In the Remove column, select the check box for the element you'd like to remove. Choose as many elements as you like. Then click the **Update List** button. The selected element will be removed from that group.

---

 **BioExtract Server**  
data access, analysis, storage, and workflow creation

Send us: [feedback!](#)  
Welcome, [clushbou@gmail.com](#) [ sign out ]

Query Extracts Tools Workflows **Groups** Help

**Demo Workflow**

**Owned Groups**  
BRIN  
✓ Sanford Research  
USD Bioteam

**Member Groups**  
Place Holder  
Users

**Additional Actions**  
Create Group

Sanford Research Members Tools Workflows Extracts

<< first < prev 1 next > last >>

Shared Tools	Description	Remove
bget	<a href="http://soap.genome.jp/">http://soap.genome.jp/</a>	<input checked="" type="checkbox"/>
embosdata	Find and retrieve EMBOSS data files	<input type="checkbox"/>
emma	Multiple sequence alignment (ClustalW wrapper)	<input type="checkbox"/>

<< first < prev 1 next > last >>

Add Elements Update List

## Inviting Members to Join a Group

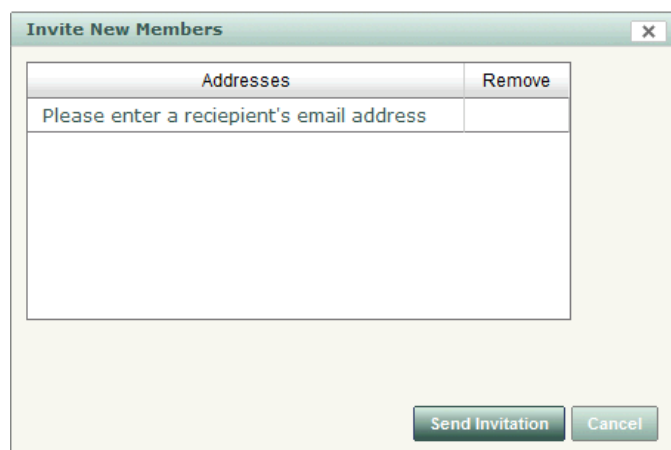
To invite a new user to a group, the group owner sends an invitation to that user. Once the recipient accepts the invitation, they will have access to the tools, workflows and data extracts owned by the group.

### To invite other people to your group:

1. Sign in to the BioExtract Server and click the Groups tab.
2. In the left panel, under the **Owned Groups** heading, select the desired group name. The group form opens in the right panel.
3. Click the **Members** tab. The Invite Members form displays.



4. Click the **Invite Members** button. An Invite New Members dialog box appears.



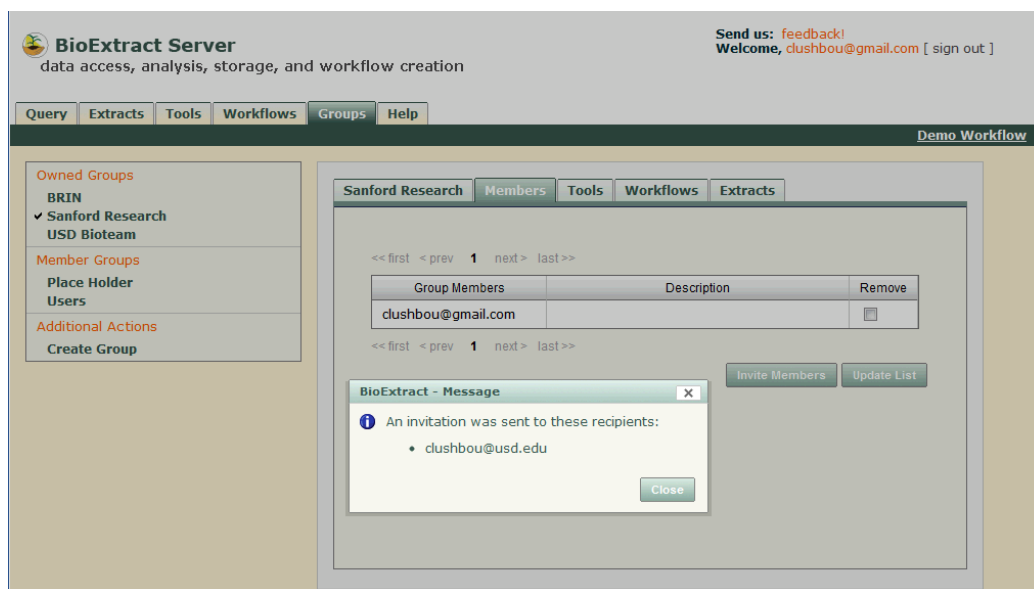
Addresses	Remove
Please enter a recipient's email address	
<div></div>	

Send Invitation Cancel

5. Select "Please enter a recipient's email address". A text box will appear.

6. Enter the email address of whomever you want to invite to the group, click the **Save** button, then click the **Send Invitation** button to complete the process. A message appears stating the invitation was sent.

When you invite people to join a group, we will immediately send email invitations to the addresses you provided. Once the recipient accepts the invitation, they will have access to the tools, data extracts and workflows owned by that group.



**To accept an invitation to join a group:**

1. If you don't have a BioExtract Server account, you'll need to create one before accepting the invitation. Open [bioextract.org](http://bioextract.org). On the top right corner of any BioExtract Server page, select **register**. Fill in the required fields and click **Add User**. Your account will be created immediately.
2. Return to the invitation email. The invitation email tells you that a BioExtract Server group wants to add you as a member. It gives you the name of the group that you've been invited to join and gives you a link to accept the invitation. Take note of the group's name for future reference.

**Collaboration Invitation from BioExtract.org**

bioextract@usd.edu

Sent: Thu 6/30/2011 8:24 AM

To: Lushbough, Carol

Hello,

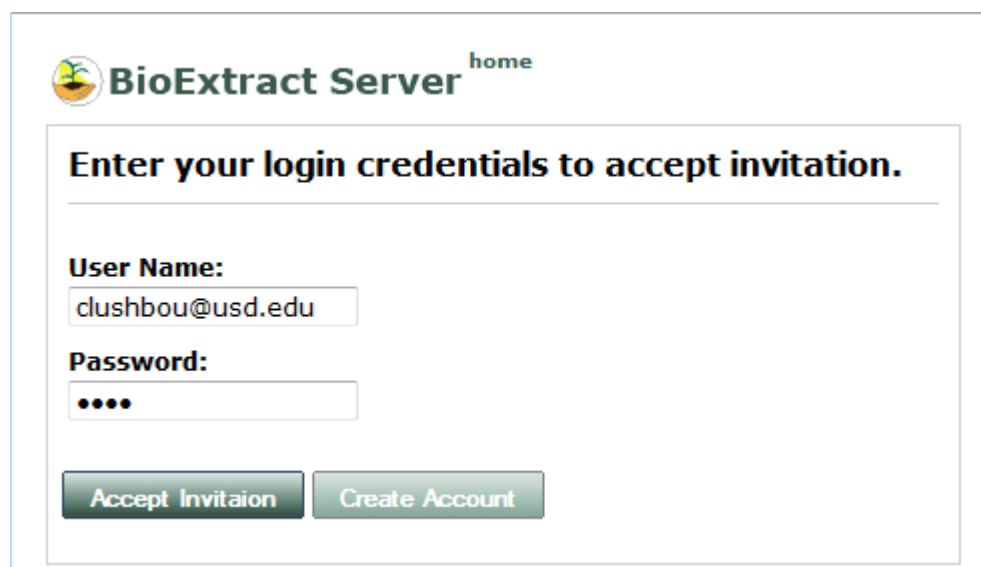
The *Sanford Research* collaboration group at [bioextract.org](http://bioextract.org) wants to add you as a member.

To accept this invitation, go to:

[http://bioextract.org/sharing/accept\\_invitation.jsp?invitationId=fef1d3a0-a939-4499-8eb1-2935010b2892](http://bioextract.org/sharing/accept_invitation.jsp?invitationId=fef1d3a0-a939-4499-8eb1-2935010b2892)

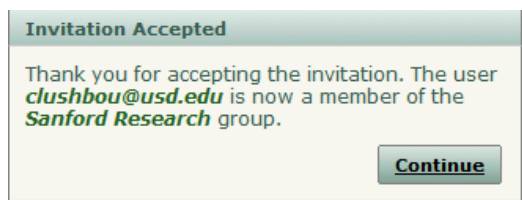
Thank you for your time.

3. Click the **Invitation** link. A BioExtract Server login window opens. Enter your BioExtract Server user name and password.



The screenshot shows the BioExtract Server login interface. At the top left is the BioExtract Server logo with a small 'home' link next to it. Below the logo is a box titled 'Enter your login credentials to accept invitation.' Inside this box, there are two input fields: 'User Name:' with the value 'clushbou@usd.edu' and 'Password:' with four dots. At the bottom of the box are two buttons: 'Accept Invitaion' and 'Create Account'.

4. Click the **Accept Invitation** button. A window displays "Thank you for accepting the invitation."



5. Click the **Continue** button. The BioExtract Server web site opens. Sign in by clicking the sign in link on the top right corner of any BioExtract Server page.
6. Click the **Groups** tab. Under Member Groups you'll see the name of the group you joined. Select the **group name**. The group interface opens in the right panel.
7. Click the **Members** tab. In the Group Members list you'll see your email address. You now have access to the tools, data extracts and workflows owned by that group.

## WORKFLOWS THROUGH MYEXPERIMENT

### About myExperiment



(<http://www.myexperiment.org>) is a collaborative environment where scientists can publish their workflows and experiment plans, share them with groups and find those of others. Workflows, other digital objects and bundles (called Packs) can be swapped, sorted and searched like photos and videos on the Web.

### Importing a BioExtract Server Workflow into myExperiment

To import a BioExtract Server Workflow into myExperiment:

1. Click the **Workflows** tab and select the desired workflow. The workflow graph and its control buttons opens in the right panel.

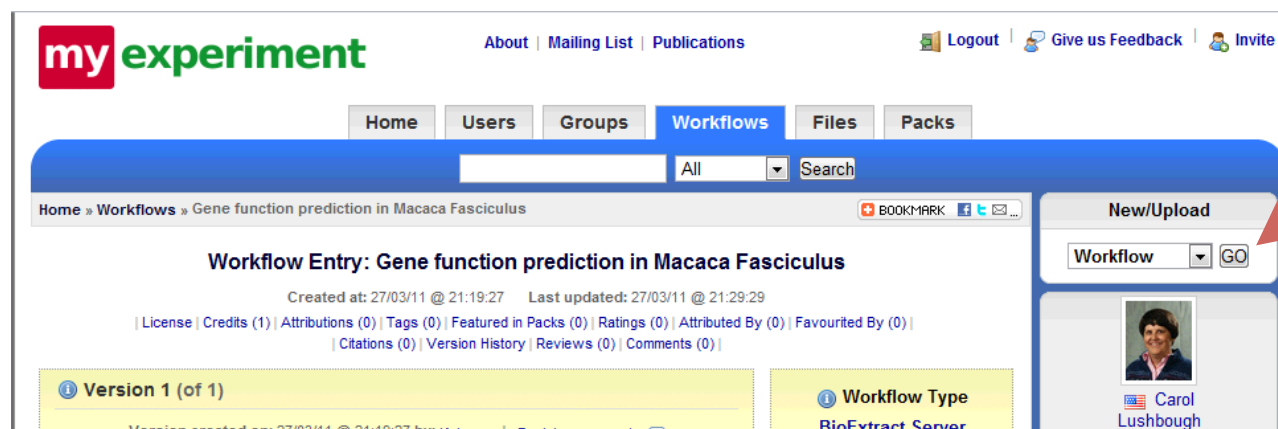
The screenshot displays the BioExtract Server web interface. At the top, the header includes the BioExtract Server logo, the tagline "data access, analysis, storage, and workflow creation", and a "Send us: feedback!" link with the email "clushbou@gmail.com" and a "sign out" button. Below the header is a navigation bar with tabs: "Query", "Extracts", "Tools", "Workflows", "Groups", and "Help". The "Workflows" tab is selected. On the left, a "Workflows" panel lists various workflows, including "Alignment", "NP\_000135 Multiple Sequence Alignment", "Nucleotide InterProScan", "Phylogenetic analysis from a gene query", "Protein alignment from a gene query", "SH2 domain (in homo) screen", "Sequence Format Conversion", "Test Copy", "Unique Asparagus petb nucleotides", "mRNA Markup", "phylogenetic analysis", "translation of dna sequence.", "tblastx non-redundant alignment", "Execute query - ( 4830 )", "tblastx - ( 4831 )", "xmknr - ( 4832 )", "FormatConversion - ( 4833 )", "FetchTranslation - ( 4834 )", "ClustalW2 - ( 4835 )", and "TCoffee - ( 4836 )". The "tblastx non-redundant alignment" workflow is selected. On the right, a "Demo Workflow" panel shows a workflow graph titled "tblastx non-redundant alignment (idle)". The graph consists of a vertical sequence of nodes: 4830, 4831, 4832, 4833, and 4834. Node 4834 branches into two nodes: 4836 and 4835. At the bottom of the workflow panel are buttons for "Details", "Execute", "Export", and "Delete".

2. Click the **Export** button. The Open dialog box opens. Save the file to your computer desktop.



3. Open the myExperiment website (<http://www.myexperiment.org>).

4. Click the Workflows tab. Click the **GO** button in the upper right corner under the **New/Upload** heading.



5. Click the **Choose File** button and select the file representing the workflow you exported from the BioExtract Server.

## Upload Workflow

### 1. Workflow file/script

tblastxnonr...ignment.xml

### 2. Main metadata

☒ Attempt to infer metadata (and possibly generate preview images) from the workflow file/script

☐ Enter custom metadata

6. Click the  button at the bottom of the screen.

7. Select the desired tags and click the **Finish** button. The workflow has been uploaded to myExperiment.

The screenshot shows the myExperiment website interface. At the top, there's a navigation bar with the myExperiment logo, links for 'About', 'Mailing List', and 'Publications', and user options like 'Logout', 'Give us Feedback', and 'Invite'. Below this is a secondary navigation bar with tabs for 'Home', 'Users', 'Groups', 'Workflows' (which is active), 'Files', and 'Packs'. A search bar is located below the tabs. The main content area is titled 'Home » Workflows' and features a green success message: 'Workflow was successfully created.' Below this, there's a section for 'Extra workflow metadata' with a 'Tags' input field. Underneath, 'Automatic tag suggestions' are listed: 'alignment', 'for', 'gene', 'genomic', 'in', 'of', 'point', 'proteins', 'sequence', 'species', and 'workflow'. A note explains that these suggestions are based on matching words in the workflow with other tags already used. A 'Summary' section follows, stating: 'You are about tag this workflow with: alignment, gene, genomic, proteins, sequence, species and workflow.' At the bottom of the main content area, there's a 'Finish' button and the text 'Complete the upload process.' On the right side, there's a sidebar with a 'New/Upload' section containing a 'Workflow' dropdown and a 'GO' button. Below this is a user profile for 'Carol Lushbough' with links to 'My Profile', 'My Messages', 'My Memberships', 'My History', and 'My News'. Further down is a 'My Stuff' section showing '1 Friend', '3 Groups', '1 Packs', '1 Files', and '6 Workflows'. The sidebar also includes sections for 'Friends' (listing 'Yuwei Lin') and 'Groups' (listing 'BioExtract Server...', 'PlantCollaborative', and 'Biomoby').